


ORIGINAL RESEARCH PAPER

Certification of the efficient random number generation technique based on single-photon detector arrays and time-to-digital converters

Andrea Stanco¹  | Davide G. Marangon^{1,2} | Giuseppe Vallone^{1,3} | Samuel Burri⁴ | Edoardo Charbon⁴ | Paolo Villoresi¹

¹Dipartimento di Ingegneria dell'Informazione, Università Degli Studi di Padova, Padova, Italy

²Now at Toshiba Europe Ltd., Cambridge Research Laboratory, Cambridge, UK

³Dipartimento di Fisica e Astronomia, Università Degli Studi di Padova, Padova, Italy

⁴Ecole Polytechnique Fédérale de Lausanne (EPFL), Neuchâtel, Switzerland

Correspondence

Andrea Stanco, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova, Via VIII Febbraio, 2, Padova 35122, Italy.
Email: andrea.stanco@unipd.it

Funding information

Ministero dell'Istruzione, dell'Università e della Ricerca, Grant/Award Number: Law 232/2016

Abstract

True random number generators (TRNGs) allow the generation of true random bit sequences, guaranteeing the unpredictability and perfect balancing of the generated values. TRNGs can be realised from the sampling of quantum phenomena, for instance, the detection of single photons. Here, a recently proposed technique, which implements a quantum random number generator (QRNG) out of a device that was realised for a different scope, is further analysed and certified [1]. The combination of a CMOS single-photon avalanche diode (SPAD) array, a high-resolution time-to-digital converter (TDC) implemented on a field programmable gate array (FPGA), the exploitation of a single-photon temporal degree of freedom, and an unbiased procedure provided by H. Zhou and J. Bruck [2, 3] allows the generation of true random bits with a high bitrate in a compact and easy-to-calibrate device. Indeed, the use of the 'Zhou–Bruck' method allows the removal of any correlation from the binary representation of decimal data. This perfectly fits with the usage of a device with non-idealities like SPAD's afterpulses, pixel cross-correlation, and time-to-digital converter non-uniform conversion. In this work, an in-depth analysis and certification of the technique presented in [1] is provided by processing the data with the NIST suite tests in order to prove the effectiveness and validity of this approach.

1 | INTRODUCTION

Quantum random number generators (QRNGs) represent one of the top-notch solutions for generating true randomness, thanks to the laws of quantum mechanics. QRNGs are used in a wide variety of applications *in primis* cryptography, which bounds this technology with quantum key distribution (QKD). Indeed, a trustful QKD system requires an appropriate QRNG device. Thus, the capability to realise a high bitrate, easy calibration, and compact QRNG devices is essential for the spreading of the QKD technology.

In this work, we present an analysis on a recently introduced QRNG technique [1], based on a single-photon temporal degree of freedom, in order to provide full assurance over the true randomness of the output bitstream. In addition

to a more detailed description of the whole system and procedure, we present the result of the data processed with the NIST suite test, which reinforces the suitability of the method. The proposed QRNG device is based on a discrete variable protocol (DV-QRNG) and falls under the so-called 'trusted device' category where full trust on the system is assumed. This follows from the fact that a possible attacker does not have physical access to the system and also from the fact that the system itself was fully characterised. This assumption bounds the security of the device, which can be considered secure only in this particular condition. A practical example includes all the applications where an external attacker does not have physical access to the device. Device-independent (DI) and semi-device-independent (Semi-DI) QRNGs offer a higher level of security since they do not (completely) rely on

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Quantum Communication* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

the device trustness [4–9], and they can also offer higher generation bitrate by exploiting continuous variables' implementation (CV-QRNG) [10, 11]. Nevertheless, with respect to a trusted device, they require more complex setups and calibration. This may also lead to higher power consumption and poor robustness against radiation effects, which are relevant features for satellite quantum communication. Indeed, a trusted QRNG represents a more secure solution with respect to a pseudo random number generator (PRNG) and classical TRNG, and a high-performance trusted one can find a place in all those applications that prefer setup simplicity, speed and cost-effectiveness over a device's side information vulnerability. Indeed, low-budget satellite QKD transmitters could be one of these [12, 13].

This work is structured as follows: in, the QRNG setup and the working principle are presented; in Section 3, the features of the single-photon CMOS sensor called LinoSPAD are described; in Section 4, a brief description of the time-to-digital converter (TDC) architecture is given; in Section 5, the Zhou–Bruck method is presented and in Section 6, the NIST test results are presented.

2 | SETUP

The system setup idea is very simple, as shown in Figure 1. It comprises a LinoSPAD device along with a neutral density (ND) filter to attenuate environment light to the single-photon level and an external computer for data processing. Single photons are detected by the CMOS single-photon avalanche diode (SPAD)

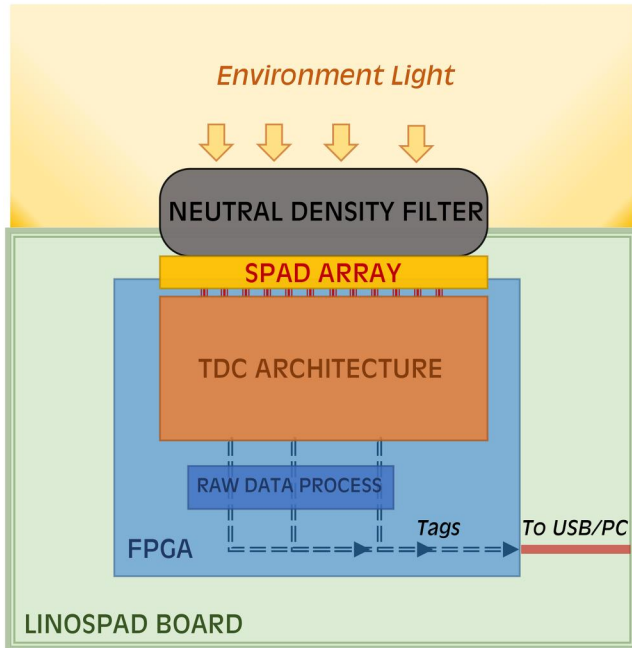


FIGURE 1 The setup schematic, which includes the LinoSPAD board and the neutral density filter. The CMOS SPAD array is directly connected to the FPGA and its TDC architecture. The random source is the environment light in the single-photon regime

array and turned into electrical signals that are directed towards the field programmable gate array (FPGA) with its implemented TDCs. The detection data are then transferred to a computer and post-processed. The only required calibration is the setting of the average photon count rate (1250 kcount/s) to prevent internal data buffer saturation. As explained in [1], due to the combination of SPAD non-idealities and TDC ones, an optimal post-processing approach is a *divide and conquer* one. Therefore, the tagging information coming from the pure sampling process provided by the 400 MHz clock is treated separately from the TDC data stream. Then, a specific post-processing procedure based on the Peres algorithm [14] is applied to the two streams. Here, we consider just the TDC data stream since it offers higher contribution to the bitrate and requires a more precise post-processing and proper randomness certification.

3 | THE LinoSPAD device

LinoSPAD is a CMOS SPAD array that was recently introduced by the AQUA laboratory [15, 16]. Its sensor is composed of an array of 256×1 pixels that are directly connected to an FPGA chip. A key feature is the TDC functionality, which was designed into the FPGA and allows the increment of the temporal resolution of the time tag a detection event is associated with. 64 TDCs implemented on the FPGA were designed to monotonically map a 2.5 ns time period to 140 codes, resulting in a nominal bin size of $\tau_{res} \approx 17.9$ ps. This resolution is achieved by sampling a delay line of 35 carry elements of 4 bits each with a 400 MHz clock. On every clock cycle, the delay line outputs a decimal number $b \in \{0, 139\}$. Thus, the final nominal bin size is given by $\tau_{res} = 1/(400\text{MHz} \times 140)$ while, as derived from the analysis in the following section, the actual bin size was evaluated to have an average value of ≈ 20 ps.

4 | TIME-TO-DIGITAL CONVERTER

Time-to-digital converters (TDCs) represent a key technology for several applications. Their key feature is the possibility to identify a certain electrical event with a very high temporal resolution usually in the range of ps. Indeed, this capability brings a huge amount of information, which can be exploited for specific applications. In our case, the more the information we get, the more the final bitrate is. Thus, exploiting TDCs hugely improves the QRNG bitrate with respect to a clock-based sampling technique (as in the ‘Coarse’ part of [1]).

The basic working principle of a TDC is to propagate a signal through a chain of logical blocks with a known propagation time, which switches one of their binary outputs as the signal passes through. By accessing the array of binary outputs, it is possible to check how long the signal is propagated through the chain and thus associate it with a precise time tag. Generally speaking, a TDC can be realised through a chain of full adders (FAs) and can be implemented on an FPGA chip, which allows great system flexibility [17–19]. Every FA is set in order to switch one of its

result outputs whenever the event reaches the FA input. By using a main system clock—usually in the hundred-MHz range—to access the array of result outputs, it is possible to check how many FAs had switched and thus how long the signal propagated. As a matter of fact, this information represents the precise time in which the event was detected within a clock cycle. Indeed, since electronic propagation times are in the ps-scale, a standard TDC resolution can vary between 1 and 100 ps.

Nevertheless, such technology was developed from components that were not meant to be used for time tagging purposes. In fact, FA chains usually include a fast carry propagation technology, which allows to speed up the computation process, that is the function they are meant to implement. Furthermore, they are not optimised to have a linear propagation time, which may vary among blocks. These features add non-linearities to the TDC. In our case, the distribution of TDC tags within a 2.5 ns clock cycle is not uniform and some of the values between 0 and 139 are not present at all (missing codes). Figure 2 shows the differential non-linearity (DNL) and the integral non-linearity (INL) of all 64 TDCs. DNL and INL are evaluated with the following formulas [20]:

$$\text{DNL}_i = \frac{n_i - n_s}{n_s} \quad (1)$$

$$\text{INL}_j = \sum_{i=1}^j \text{DNL}_i \quad (2)$$

where n_s is equal to the total number of detected events divided by 140 and n_i corresponds to the number of detected events in the i th bin. To avoid missing events, the TDCs sample a carry chain with a propagation time longer than the sampling period, resulting in unused codes and a large excursion in the INL. Each of the four output blocks of the chain typically has

two fast and two slow outputs resulting in a characteristic pattern in the DNL plot. The output encoding ensures code monotonicity and the maximum possible resolution. Depending on the application, calibration or post-processing of the TDC codes is necessary. The plots show a non-uniform distribution where some values are more likely than others and there are no events below ≈ 15 . Furthermore, at least 8 bits are needed to describe 140 different values, which worsen the distribution even more since values between 139 and 255 are not present. However, the autocorrelation plot in Figure 3 shows that the serial correlation of the decimal codes is within the fiducial limits, which assures the possibility of generating quantum random numbers by exploiting the temporal degree of freedom of the single-photon and Peres algorithm. The only requirement is to eliminate non-uniform distribution of decimal codes before applying the Peres algorithm. For this reason, the Zhou–Bruck algorithm is applied. It is worth stressing that even if the time resolution can be considered state-of-the-art, the non-linearity could clearly be improved. However, this work does not deal with TDC time-domain optimisation but rather highlights the robustness of the randomness extraction method without any modification of the entropy source or custom device calibration.

5 | THE ‘ZHOU–BRUCK’ METHOD

The Zhou–Bruck algorithm was introduced in [2] and later analysed in [3]. It allows the safe application of the Peres [14] algorithm to a biased dice. In our case, it perfectly serves as a procedure for applying the Peres algorithm to a correlated bitstring. As a matter of fact, the Peres algorithm only works to remove the bias and it cannot be applied to a correlated bitstring since it will not remove correlation. According to the plots in Figure 2, presence of a non-uniform distribution is

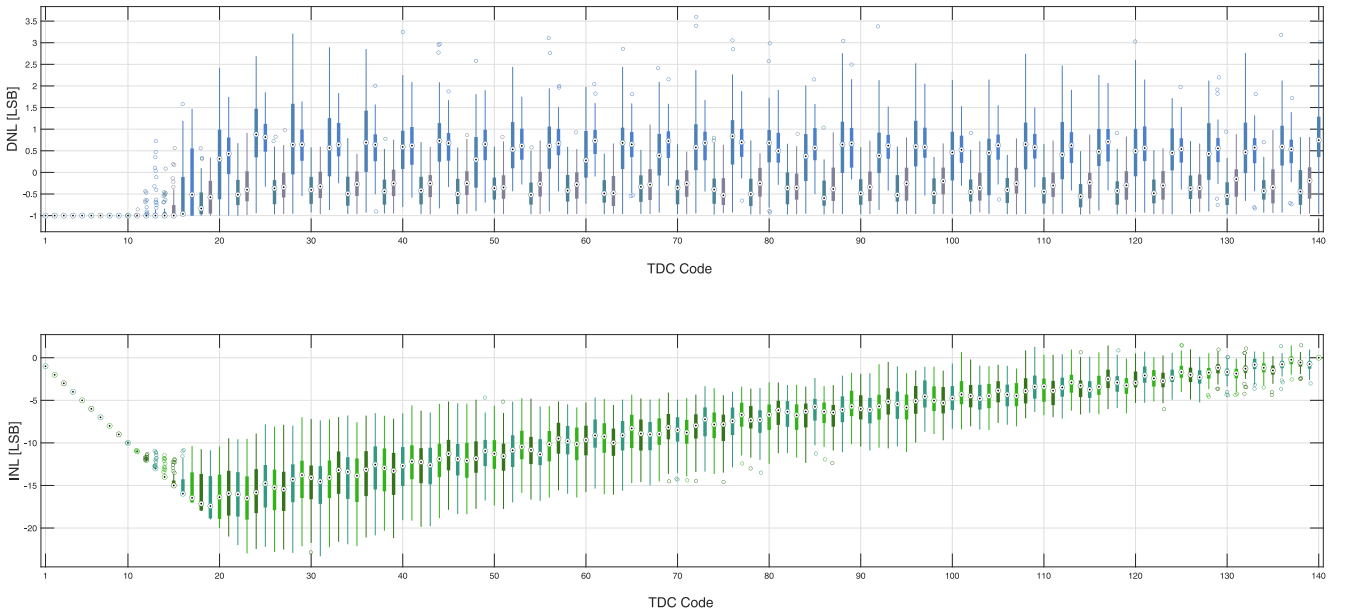


FIGURE 2 Differential non-linearity (DNL) and integral non-linearity (INL) boxplots for all the 64 TDCs. To avoid missing events, the delay lines are longer than the sampling period, which results in unused codes and a large negative excursion of the INL.

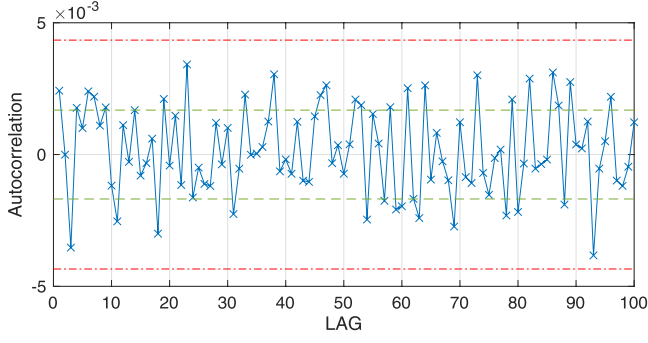


FIGURE 3 Autocorrelation plot for the events of pixel 1 evaluated before applying the Zhou–Bruck method, that is by considering the codes in their decimal form between 0 and 139. The serial autocorrelation re-enters within the limit of acceptance. Green dashed lines represent standard deviation while red dot-dashed lines are 99% confidence limits

clear. Conversion from decimal to a binary base, which is required to apply the Peres algorithm, introduces a correlation among bits (this can be verified in a heuristic way). Figure 4 shows the working principle of the method. By taking a stream of numbers in their binary representation with length L and organising them by columns $C_1 \rightarrow C_L$, it is possible to group the bits in specific uncorrelated sub-sets $X(1) \rightarrow X(2^L - 1)$ and apply the Peres algorithm individually to each set. The rule to populate the first seven sub-sets is the following:

- $X(1)$ is equal to the first column C_1 .
- $X(2)$ contains the bits of C_2 for those rows of C_1 containing a 0.
- $X(3)$ contains the bits of C_2 for those rows of C_1 containing a 1.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	...	C_L
R_1	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	b_{17}	b_{18}		b_{1L}
R_2	b_{21}	b_{22}	b_{23}	b_{24}	b_{25}	b_{26}	b_{27}	b_{28}	...	b_{2L}
...			
R_N	b_{N1}	b_{N2}	b_{N3}	b_{N4}	b_{N5}	b_{N6}	b_{N7}	b_{N8}		b_{NL}

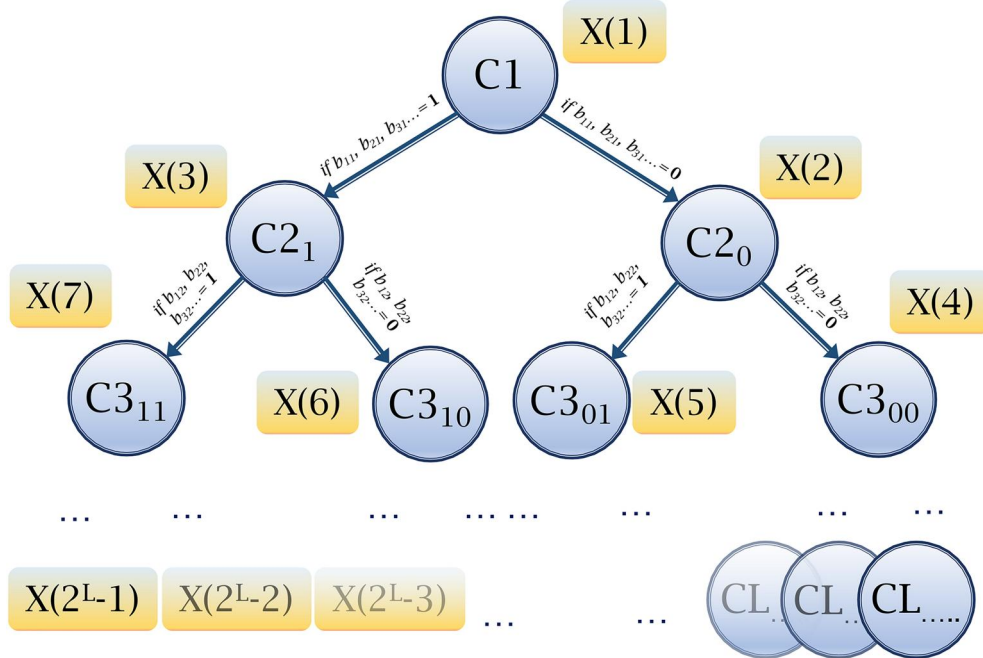


FIGURE 4 The Zhou–Bruck method example table (top) and tree graph (bottom). The table shows a possible stream of N numbers represented in their binary form with length L . The graph shows how to group the bits from the table. Every set is created by selecting a specific bit of a column according to the value of the bits on the same row of previous columns. For instance, set $X(2)$, built using bits of C_2 , can contain b_{12} only if $b_{11} = 0$, b_{22} only if $b_{21} = 0$ etc. On the contrary, set $X(3)$, built using bits of C_2 , can contain b_{12} only if $b_{11} = 1$, b_{22} only if $b_{21} = 1$ etc.; set $X(4)$, built using bits of C_3 , can contain b_{13} only if $b_{11}b_{12} = 00$, b_{23} only if $b_{21}b_{22} = 01$ etc. and set $X(5)$, built using bits of C_3 , can contain b_{13} only if $b_{11}b_{12} = 01$, b_{23} only if $b_{21}b_{22} = 01$ etc

TABLE 1 Result of the NIST test suite on the pre-processed data

Test's name	p -value	Proportion	Result
Frequency	0.000000	38/500	FAILED
Block Frequency	0.000000	57/500	FAILED
Cumulative Sums	0.000000	2/500	FAILED
Cumulative Sums	0.000000	2/500	FAILED
Runs	0.000000	47/500	FAILED
Longest Run	0.000000	310/500	FAILED
Rank	0.079538	499/500	PASSED
FFT	0.000000	403/500	FAILED
Non-Overlapping Template	0.000000	23/500	FAILED
Overlapping Template	0.000000	186/500	FAILED
Universal	0.000000	234/500	FAILED
Approximate Entropy	0.000000	126/500	FAILED
Random Excursions	—	9/9	FAILED
Random Excursions Variant	—	9/9	FAILED
Serial	0.000000	273/500	FAILED
Linear Complexity	0.645448	497/500	PASSED

Note: The minimum pass rate for random excursions and random excursions' variant tests is 8/9 while for all the other tests is 488/500.

- $X(4)$ contains the bits of C_3 for those rows of C_1 and C_2 containing 00.
- $X(5)$ contains the bits of C_3 for those rows of C_1 and C_2 containing 01.
- $X(6)$ contains the bits of C_3 for those rows of C_1 and C_2 containing 10.
- $X(7)$ contains the bits of C_3 for those rows of C_1 and C_2 containing 11.

The rule can be expanded till the creation of the $X(2^L - 1)$ set. According to Figure 4, by following this rule it is clear that a tree structure is created until $2^L - 1$ sets are created. Thanks to this grouping, each subset contains uncorrelated bistrings, which can then be safely processed with the Peres algorithm to finally remove the binary biasing.

6 | RESULTS

We applied the Zhou–Bruck method to a data set of 723.408376 Mbits, corresponding to 90.426047 Mtags acquired in ≈ 2.56 s, and generated a final random sequence of 571.618824 Mbits with an equivalent bitrate of ≈ 223 Mbit/s and overall extraction efficiency above 90%. We performed the NIST test suite on both the data collections. The pre-processed data set failed most of the tests while the post-processed one successfully passed all the tests, certifying the validity of the described method and the capability of the Zhou–Bruck algorithm to remove any correlation even from a

TABLE 2 Result of the NIST test suite on the post-processed data

Test's name	p -value	Proportion	Result
Frequency	0.078567	494/500	PASSED
Block Frequency	0.316052	494/500	PASSED
Cumulative Sums	0.366918	495/500	PASSED
Cumulative Sums	0.790621	495/500	PASSED
Runs	0.459717	499/500	PASSED
Longest Run	0.076658	494/500	PASSED
Rank	0.624627	492/500	PASSED
FFT	0.950247	495/500	PASSED
Non-Overlapping Template	0.003551	497/500	PASSED
Overlapping Template	0.950247	495/500	PASSED
Universal	0.411840	495/500	PASSED
Approximate Entropy	0.858002	494/500	PASSED
Random Excursions	0.131806	311/318	PASSED
Random Excursions Variant	0.024383	314/318	PASSED
Serial	0.222480	497/500	PASSED
Linear Complexity	0.978072	497/500	PASSED

Note: The minimum pass rate for random excursions and random excursions' variant tests is 309/318 while for all the other tests is 488/500.

highly correlated bitstring affected from several system non-idealities. Tables 1 and 2 show the test results. For the sake of simplicity, for those tests with multiple results, the one with the worse p -value is reported.

7 | CONCLUSION

In this work, we analysed and tested a quantum random number generator capable of producing certified random numbers with a bitrate equal to 223 Mbit/s starting from a photon count rate equal to 1250 kcount/s, a time resolution of 17.9 ps, a 400 MHz sampling clock, and a 256 CMOS SPAD array. The QRNG uses environment light as the photon source and an ND filter for the single-photon regime. The generated bitstring was tested with the NIST suite and passed all the tests. The same test was conducted before the application of the post-processing approach and all the tests failed. This result shows the feasibility of high-performance QRNGs with simple setup and calibration, compact form factor, and exploitation of environment light. It also confirms the validity of the Zhou–Bruck method and its strong uncorrelation power. Future steps will focus on implementing the post-processing approach directly on the FPGA without needing an external PC and improve the device compactness and usability.

ACKNOWLEDGMENTS

Part of this work was supported by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) (the Italian

Ministry of Education, University and Research) under the initiative ‘Departments of Excellence’ (Law 232/2016). A. Stanco would like to thank Dr. Avesani for useful hints in setting the NIST suite.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Andrea Stanco  <https://orcid.org/0000-0001-8442-9055>

REFERENCES

1. Stanco, A., et al.: Efficient random number generation techniques for cmos single-photon avalanche diode array exploiting fast time tagging units. *Phys. Rev. Res.* 2, 023287 (2020)
2. Zhou, H., Bruck, J.: A Universal Scheme for Transforming Binary Algorithms to Generate Random Bits from Loaded Dice. *arXiv:12090726* (2012). <https://arxiv.org/abs/1209.0726>
3. Pae, S.: Binarization trees and random number generation. *IEEE Trans. Inf. Theor.* 66(4), 2581–2587 (2020)
4. Pironio, S., et al.: Random numbers certified by bell’s theorem. *Nature.* 464, 1021–1024 (2010)
5. Christensen, B.G., et al.: Detection-loophole-free test of quantum non-locality, and applications. *Phys. Rev. Lett.* 111, 130406 (2013)
6. Bierhorst, P., et al.: Experimentally generated randomness certified by the impossibility of superluminal signals. *Nature.* 556(7700), 223–226 (2018)
7. Liu, Y., et al.: High-speed device-independent quantum random number generation without a detection loophole. *Phys. Rev. Lett.* 120, 010503 (2018)
8. Gómez, S., et al.: Experimental nonlocality-based randomness generation with nonprojective measurements. *Phys. Rev. A.* 97, 040102 (2018)
9. Ma, X., et al.: Quantum random number generation. *npj Quant. Inf.* 2, 16021 (2016)
10. Marangon, D.G., Vallone, G., Villoresi, P.: Source-device-independent ultrafast quantum random number generation. *Phys. Rev. Lett.* 118, 060503 (2017)
11. Avesani, M., et al.: Source-device-independent heterodyne-based quantum random number generator at 17 gbps. *Nat. Commun.* 9(1), 5365, (2018)
12. Oi, D.K., et al.: Cubesat quantum communications mission. *EPJ Quant. Technol.* 4(1), 6 (2017)
13. Balossino, A., et al.: SeQBO - A miniaturized system for quantum key distribution. In: *Proceedings of the International Astronautical Congress, IACVolume 2020-October 2020 71st International Astronautical Congress, IAC 2020, Virtual, Online, 12 October 2020 - 14 October 2020*, 166680. <http://iafastro.directory/iac/paper/id/59867/summary/>
14. Peres, Y.: Iterating von neumann’s procedure for extracting random bits. *Ann Statist.* 20(1), 590–597 (1992)
15. Burri, S., et al.: A time-resolved 256x1 CMOS SPAD line sensor system featuring 64 FPGA-based TDC channels running at up to 8.5 giga-events per second. *Optical Sensing and Detection IV*, vol. 9899, pp. 57–66. International Society for Optics and PhotonicsSPIE (2016)
16. Burri, S., Bruschini, C., Charbon, E.: Linospad: A compact linear spad camera system with 64 fpga-based tdc modules for versatile 50 ps resolution time-resolved imaging. *Instruments.* 1(1) (2017)
17. Favi, C., Charbon, E.: A 17 ps time-to-digital converter implemented in 65 nm fpga technology. *Proceedings of the ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, pp. 113–120. *FPGA ’09.ACM*, New York (2009)
18. Fishburn, M., et al.: A 19.6 ps, fpga-based tdc with multiple channels for open source applications. *IEEE Trans. Nucl. Sci.* 60(3), 2203–2208 (2013)
19. Song, J., An, Q., Liu, S.: A high-resolution time-to-digital converter implemented in field-programmable-gate-arrays. *IEEE Trans. Nucl. Sci.* 53(1), 236–241 (2006)
20. Kalisz, J.: Review of methods for time interval measurements with picosecond resolution. *Metrologia.* 41(1), 17–32, (2003)

How to cite this article: Stanco, A., et al.: Certification of the efficient random number generation technique based on single-photon detector arrays and time-to-digital converters. *IET Quant. Comm.* 1–6 (2021). <https://doi.org/10.1049/qtc2.12018>