Jailbreak Imagers: Transforming a Single-Photon Image Sensor into a True Random Number Generator.

Samuel Burri[1], Damien Stucki[2], Yuki Maruyama[3], Claudio Bruschini[1], Edoardo Charbon[3], Francesco Regazzoni[3]
[1]EPFL, School of Engineering, Lausanne, 1015 Switzerland. Phone +41.21693.7524. E-mail: samuel.burri@epfl.ch
[2]ID Quantique, 1227 Carouge, Switzerland. Phone +41.22.3018371.
[3]Delft University of Technology, 2628 Delft, Netherlands. Phone +31.15.2783667.

Abstract—**In this paper we present a large array of SPADs and we discuss its suitability for applications different than imaging, exploring in particular how to transform it into a high speed True Random Number Generator (TRNG). The proposed matrix comprises 512x128 independent cells that convert photons into a raw bit-stream, which, as ensured by the properties of quantum physics, is characterized by a very high level of randomness. The sequences are read out in a 128-bit parallel bus, concatenated, and pipelined onto a de-biasing filter. Our results, achieved by coupling two matrices, show that our architecture can reach up to 5 Gbit/s while consuming 25pJ/bit, to our knowledge the lowest in a TRNG to date.**

I. Introduction

Random numbers are required in many applications, ranging from password or cryptographic key generation to gaming (e.g. winning number drawing or card deck shuffling). Although for certain applications pseudorandom numbers are sufficient and even desirable, true random numbers are increasingly used either for security or regulatory reasons. The emergence of quantum key distribution as a technique to enable secure key exchange according to information theory and the pervasive diffusion of privacy sensitive applications, such as web services for e-commerce and e-health, push for the development of low cost true random number generators in the multi-Mb/s range for clients and in the multi-Gb/s range for servers.

High speed True Random Number Generators (TRNGs) have been proposed based on mechanisms, such as thermally induced jitter from ring oscillators, block RAM write collisions, flip-flop metastability on FPGAs [6] and ASICs [4], etc. TRNGs may also exploit optical effects. In [5] and [2], the use of superluminescent LEDs and lasers was proposed as a source of physical entropy achieving rates of up to 300Gb/s, however, both TRNGs were implemented in non-standard processes.

An effective way to create an optical TRNG is to use the quantum nature of photons. Reference [7] for instance measured the quantum phase noise of a laser operating at low intensity levels for rates up to 6.25 Gbits/s. However, the system is fabricated in a custom process and the operating conditions to achieve stable, high-quality random numbers are hard to achieve and/or to maintain. To date, commercial quantum random generators can only reach speeds of 150Mb/s and are often built in expensive custom processes. Alternatively, CMOS quantum random number generators have been proposed by a number of authors, usually in the multi-Mb/s. CMOS quantum random number generators rely on the same design principles and techniques used for realizing images, but, usually the number of detectors used in parallel is very limited. In fact, a complete and exhaustive study of the scalability of this approach was still missing thus the use of massively parallel quantum random number generators is so far mainly unexplored.

In this paper, we bridge this gap by exploring the suitability of an imager composed of a large number of pixels as a True Random Number Generator. Random collapse of wave function in the plane X-Y of the detector is used as source of entropy. A large number of detectors, implemented using a standard CMOS technology, composes the array which is organized in a regular geometry. The detectors are used in parallel to increase the overall throughput of the TRNG. This approach is sound because each detector tends to respond independently from the others, assuming near-zero crosstalk.

Our design exploits SPADs as detectors and a LED as photon source. If properly designed, SPADs exhibit the needed low optical and electrical crosstalk, while, the bit-stream of each SPAD can be considered a random process, assuming zero afterpulsing. Afterpulsing, in fact, introduces a correlation between subsequent pulses, thus degrading the quality of the randomness in a similar way as crosstalk.

We evaluate the quality of our design as a random number generator studying the effects of detector- and source-related properties, while varying the number of activated pixels, as well as supply voltage and temperature. Finally, the throughput and the quality of the TRNG were validated using the NIST and diehard test suites.

The paper is organized as follows. Section II describes the architecture of our TRNG. Section III reports the performance of our chip.

II. Design

The overall system consists of two identical, tightly coupled cameras operating independently. Each of them, as detailed in Figure 1, comprises three main units: the photon source, the detector array, and an algorithmic post-processing unit. The photon source is a pulsed LED with peak emission at 830nm, and it is placed at the center of the array at a distance of 2cm to allow homogeneous illumination of the whole matrix.

The detector array consists of a dual 512x128 pixel array. Each pixel comprises a SPAD, implemented as in Figure 2, and a one-bit memory element. The SPAD is quenched via N1; the cathode drives N3 which in turn sets the latch formed by N4-N5-N9-N10, upon photon detection. The NMOS latch is controlled by TOPGATE that can also be used to save power,

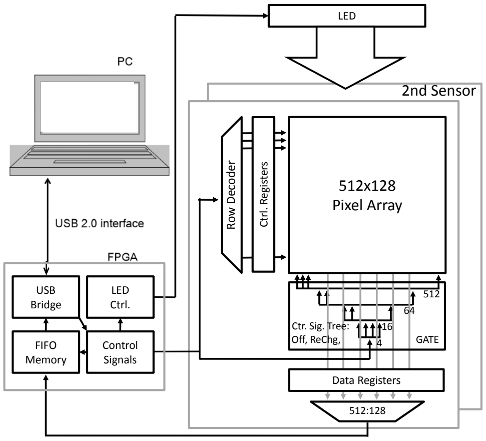


Fig. 1: Block diagram of the proposed true random number generator. The pulsed light is produced by a LED with peak emission at 830nm, which is placed on the center of the array at a distance of 2cm to allow homogeneous illumination of the whole matrix. The internal memory bank (512 memory elements) is connected with external memory elements, which are read out in parallel and concatenated to produce the bit stream. The bit-stream is input to a filter to remove the bias of the sequence and final stream is output outside.

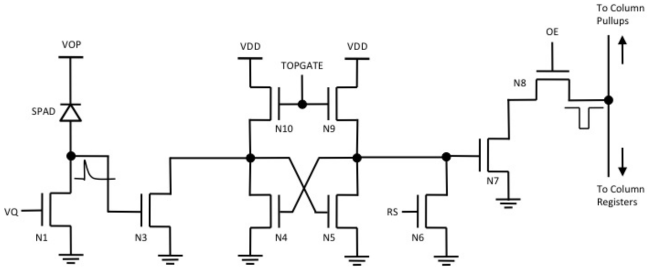


Fig. 2: Schematic of the 9T pixel in the proposed TRNG. The SPAD is quenched via N1; the anode drives N3 which in turn sets the latch formed by N4-N5-N9-N10, upon photon detection. The NMOS latch is controlled by TOPGATE that can also be pulsed to save power, and reset by RS via N6. The output of the latch controls the pulldown N7 that is used to change the column line via select transistor N8 controlled by signal OE. The latter is pulled up at the top of the column and read out at the bottom.

and reset by RS via N6. The output of the latch controls the pulldown N7 that is used to change the column line via select transistor N8 controlled by signal OE. The latter is pulled up at the top of the column and read out at the bottom.

Each pixel contains local shutter transistors for fast response and a memory where photon detections during the active time are registered. Through selection and reset transistors, a full line of the sensor will be read out and reset in one operation.

The chip supports both global-shutter and rolling shutter modes. The row decoder signal enables the desired row in the array at the time of the read-out after which the memories in the row are reset. Every column is read out independently via a fast memory and a serializer. The entire content of one array (65,536 bits) is completely read out in 6.4μs (frame duration) via a 128-bit bus; note that a bit-stream of 10.2Gb/s is achieved by each array irrespectively of the sequence and the number of rows read out in the frame duration.

To acquire one frame of random data, the memories in the sensor are reset and the SPADs are activated by applying the excess voltage which brings them in Geiger regime. The LED is then activated for a duration which will give each SPAD a 50% chance of being triggered by a photon. After deactivation of the SPAD frontend circuit, the resulting random bits are read out and the memories reset again for the next acquisition.

The bit-streams are pipelined onto the algorithmic post-processing unit, which implements a von Neumann filter to de-bias the sequence[1]. The filter reduces the throughput from a raw bit-stream of 20.4Gb/s to approximately 5Gb/s. The overall system is controlled by a dedicated FPGA which uploads the streams of bits to the PC using the USB 2.0 interface.

The chip was implemented using a standard 0.35μm CMOS technology and it measures 12.3mm x 3.3mm. The micrograph of one of the two pixel arrays is captured on Figure 3. The figure also reports the detail of the pixel used to build the complete true random number generator, and the photograph of both the pixel arrays mounted and wire-bonded on a PCB.

III. Results

In order to be suitable as a high speed TRNG, the output of the coupled imagers must fulfill several statistical properties and reach a certain throughput. Statistical properties were verified by means of the NIST test suite (Table I). Throughput was analyzed under different conditions.

We define Random Bit Efficiency (RBE) as the ratio between de-biased bit-streams and the raw bit-streams. RBE was computed for a range of temperatures from -25°C to 70°C biasing the SPADs in the pixels at an excess bias voltage from 2.0V to 5.5V. The LED was biased at an average power of 100μW and pulsed at 156kHz with a duty cycle of 0-15%, i.e. a pulse length from 0 to 900ns. The RBE is plotted in Figure 4 as a function of excess bias voltage and LED pulse length at the indicated temperature range. The plots demonstrate that an optimum is found in a large region of operation. The plot in Figure 5(a) shows the throughput of the TNRG as a function of activated pixels before and after de-biasing; with fewer pixels, the minimum readout cycle of 50ns is used. At this speed, afterpulsing degrades the quality of the TRNG sequences and thus the usable throughput after de-biasing, is relatively low. Increasing the number of activated pixels has the effect of increasing the readout cycle, thereby reducing afterpulsing and thus increasing RBE and the overall throughput. The relation between afterpulsing and readout cycle time is complex however it becomes negligible at readout cycles in the order of a few μs [1], as shown in Figure 5(b).

[1] Random sequences are usually filtered to remove potential biases of the source

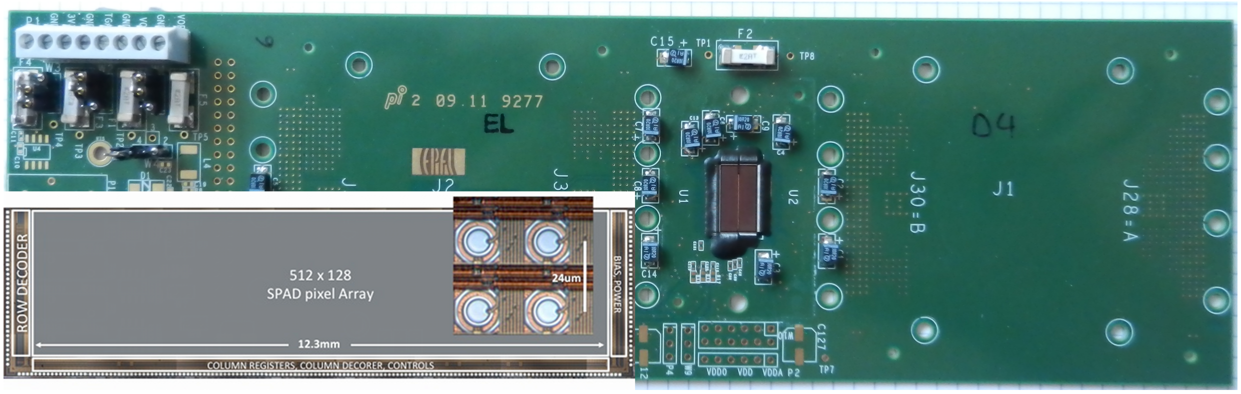


Fig. 3: Photograph of the chip mounted and wire-bonded on a PCB. The insets show the micrograph of a single marix and the detail of the pixels, which compose the complete array, respecitvely. The chip was fabricated in standard 0.35μm CMOS technology. It measures 12.3mm x 3.3mm.

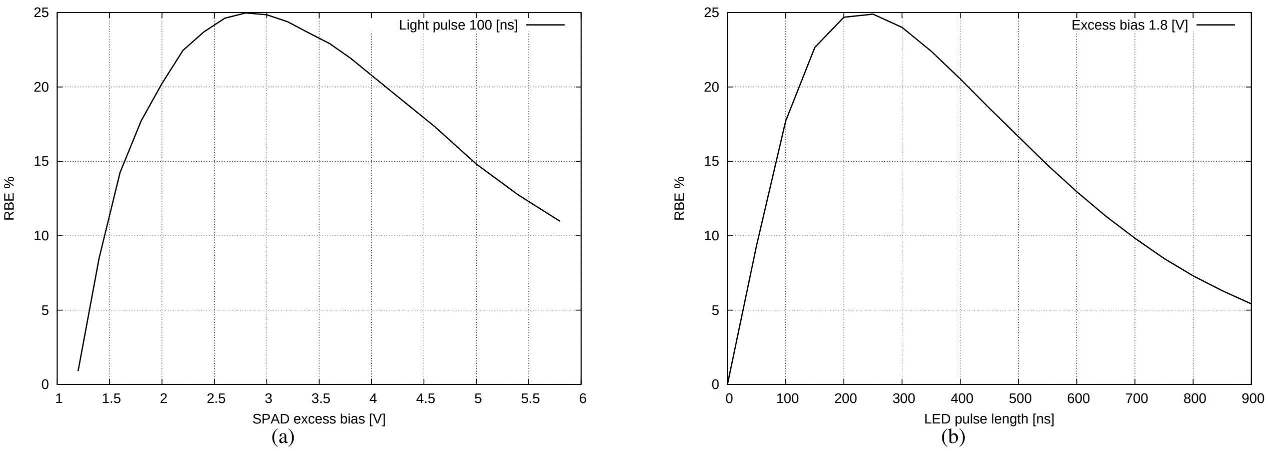


Fig. 4: Measured random bit efficiency (RBE) vs. (a) excess bias voltage and (b) LED pulse length. The plot shows a maximum efficiency of 25%. These measurements were repeated in the temperature range with no statistical deviation.

A performance comparison between the proposed TRNG and the literature is illustrated in Table II. To the best of our knowledge, the proposed TRNG is the fastest implemented in a standard CMOS process, while higher throughput is only achieved by Wei et al [5], using a non-quantum process in a custom, non-CMOS technology. The proposed TRNG has also the highest energy efficiency ever reported in any technology.

IV. Conclusion

In this paper we explored the scalability of one of the most appealing applications for quantum CMOS: true random number generation. In particular, we coupled two matrices each consisting of 512x128 independent cells that convert photons into a raw bit-stream. The experiments we conducted using standard tests for randomness proved that the sequences produced by our TRNG are characterized by a very good random properties.

The performance is measured on a device manufactured in standard CMOS process. Measurements show that our architecture can reach up to 5 Gbit/s while consuming 25pJ/bit. Our results prove the scalability and performance for any random number generators based on SPADs, while achieving the lowest power consumption to date.

Acknowledgements

The authors are grateful to Xilinx for providing the FPGAs used in the readout system. This work has been partially supported by the Swiss NCCR-QP, NCCR-QSIT, NCCR-MICS, Swiss Experiment, and EMRP (project IND06-MIQC). The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

References

[1] M. Fishburn. *Fundamentals of CMOS Single-Photon Avalanche Diodes*. PhD Thesis. TU Delft, Sep 2012.

[2] I. Kanter, Y. Aviad, I. Reidler, E. Cohen, and M. Rosenbluh. An Optical Ultrafast Random Bit Generator. *Nature Photonics*, 4:58–61, 2010.

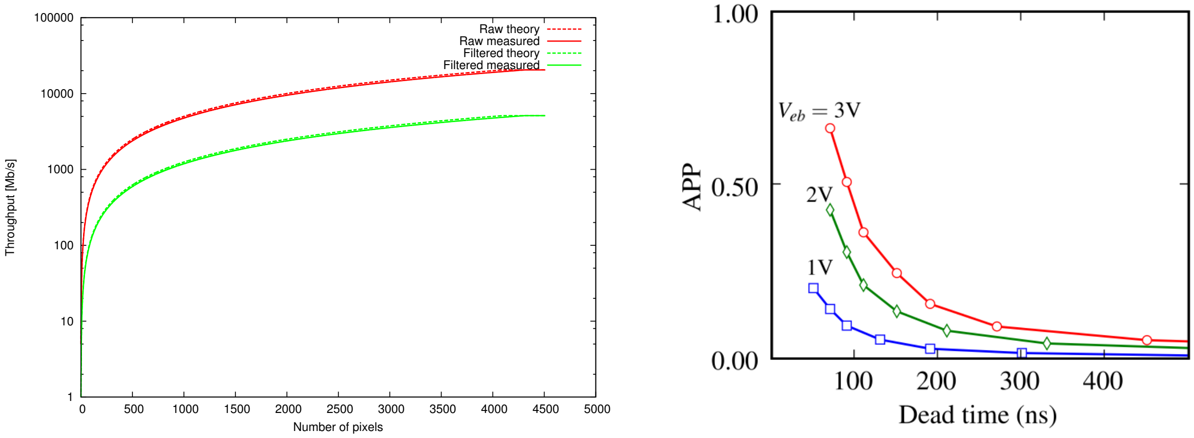
[3] M. Matsumoto, S. Yasuda, R. Ohta, K. Ikegami, T. Tanamoto, and S. Fujita. $1200\mu m^2$ Physical Random-Number Generators Based on SiN MOSFET for Secure Smart-Card Application. In *ISSCC 2008*, pages 414–415, Feb 2008.

[4] F. Pareschi, G. Setti, and R. Rovatti. Implementation and Testing of High-Speed CMOS True Random Number Generators Based in Chaotic Systems. In *IEEE Trans. Circ. & Sys.*, volume 57–I(12), pages 3124–3137, Oct 2010.

[5] W. Wei, G. Xie, A. Dang, and H. Guo. High-Speed and Bias-Free Optical Random Number Generator. *Photonics Technology Letters*, 24(6):437–439, June 2012.

[6] K. Wold and S. Petrovic. Optimizing Speed of a True Random Number Generator in FPGA by Spectral Analysis. *ICCIT*, Nov 2009.

[7] F. Xu, B. Qi, X. Ma, H. Xu, H. Zheng, and H.-K. Lo. Ultrafast Quantum Random Number Generation based on Quantum Phase Fluctuations. *Optics Express*, 20(11):12366–12377, Nov 2012.



(a)

(b)

Fig. 5: (a) Measured and theoretical throughput vs. number of pixels activated in an experiment with 2.8V of excess bias. The throughput is shown before and after de-biasing, whereas a 4x reduction is caused by von Neumann based de-biasing, as expected using this optimal excess bias. The plot shows that the theoretical linear relation between throughput and the number of active pixels is achieved with the increase of the number of active pixels. However, due to 5% of the pixels that are non-functional, a certain level of redundancy must always be added and thus the measured and theoretical curves (before and after de-biasing) do not perfectly overlap. (b) Afterpulsing probability as a function of the dead time between readout operations, equivalent to the read cycle time [1].

TABLE I: Results of the NIST tests applied to a sequence generated with a LED pulse length of 100ns and an excess bias voltage of 2.8V. The tests were run on the data from the de-biasing filter.

Test	Accept Threshold	Von Neumann	Pass / No Pass
Frequency	0.951464	0.9833	Y
BlockFrequency	0.951464	0.9833	Y
CumulativeSum	0.951464	0.9833	Y
Runs	0.951464	1.0000	Y
LongestRun	0.951464	1.0000	Y
Rank	0.951464	1.0000	Y
FFT	0.951464	0.9833	Y
NonOverlapping Template	0.951464	0.9667	Y
Universal	0.951464	1.0000	Y
ApproximateEntropy	0.951464	1.0000	Y
RandomExcursion	0.951464	0.9744	Y
RandomExcursion Variant	0.942202	0.9744	Y
Serial	0.951464	1.0000	Y
LinearComplexity	0.951464	1.0000	Y

TABLE II: Comparison of the proposed TRNG performance and the state-of-the-art. *) The area refers to the active core. **) Data not available.

Measure	Min	Typ	Max							Unit
Reference	This work			[7]	[3]	[6]	[4]	[5]	[2]	
Reported Throughput (R = Raw, P = Post-Processed)	10 (R)	15 (R)	20 (R)	6.7 (P)	0.02 (P)	0.3 (R)	0.04 (R)	280 (R)	300 (R)	Gb/s
Temp. Range	-25	27	70	**	**	**	**	**	**	°C
Vdd	3.0	3.3	3.6	N/A	N/A	N/A	N/A	N/A	N/A	V
Excess Bias	1.2	1.8	4.0	N/A	N/A	N/A	N/A	N/A	N/A	V
LED Pulse Length	50	100	500	N/A	N/A	N/A	N/A	N/A	N/A	ns
LED Duty Cycle	0.8	2	10	N/A	N/A	N/A	N/A	N/A	N/A	%
Power		500		**	1.9	**	29	**	**	mW
Area		7.7		**	0.012	**	0.752	**	**	mm²
Energy/bit		25		**	950	**	725	**	**	pJ/bit
Technology	0.35μm CMOS			Custom (InGaAs)	SiN MOSFET	CMOS (FPGA)	0.35μm CMOS	Custom	Custom	