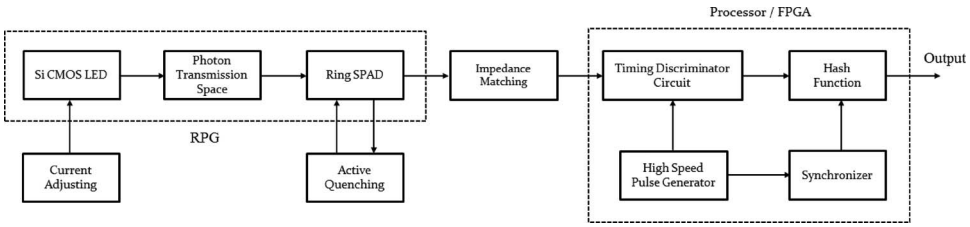
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

A Monolithic Silicon Quantum Random Number Generator Based on Measurement of Photon Detection Time

Volume 7, Number 5, October 2015

Abbas Khanmohammadi
Reinhard Enne
Michael Hofbauer
Horst Zimmermanna



DOI: 10.1109/JPHOT.2015.2479411
1943-0655

IEEE photonics SOCIETY

IEEE

A Monolithic Silicon Quantum Random Number Generator Based on Measurement of Photon Detection Time

Abbas Khanmohammadi, Reinhard Enne, Michael Hofbauer, and Horst Zimmermanna

Faculty of Electrical Engineering and Information Technology, Institute of Electrodynamics, Microwave and Circuit Engineering, Vienna University of Technology, 1040 Vienna, Austria

DOI: 10.1109/JPHOT.2015.2479411

Manuscript received June 24, 2015; accepted September 10, 2015. Date of publication September 16, 2015; date of current version October 5, 2015. This work was supported by the Vienna University of Technology Library through its Open Access Funding Program. Corresponding author: A. Khanmohammadi (e-mail: akhanmohammadi@emce.tuwien.ac.at).

Abstract: In this paper, a nondeterministic random number generator based on detection of the single photons emitted by an Si-CMOS-LED light source integrated for the first time on the detector chip is presented and experimentally demonstrated. We use a ring-shaped single-photon avalanche diode (SPAD) around the Si-CMOS-LED fabricated in 0.35-μm HV-CMOS technology to generate random events. The time intervals between single-photon events are independent quantum random variables. A field-programmable gate array (FPGA) digitizes the time variables to the stream of random bits. Bias in the raw data due to the nonuniform distribution of the time intervals is removed by postprocessing in a special configuration of XOR gates to improve the randomness of the generated random bits. The quantum random numbers in 1-Gb streams with bit generation rate of 1 Mb/s were directly delivered to a personal computer (PC) and passed all statistical tests from ENT, STS, and DIEHARD, as well as for more accuracy correlation and bias tests applied on these streams.

Index Terms: Quantum random number generator, single-photon avalanche diode, Si-CMOS-LED.

1. Introduction

Random numbers are important in wide areas of science and technology, including numerical simulations, stochastic modeling, generation of the classical and quantum cryptographic keys, random initialization of the variables in cryptographic protocols, online hazard games, communication security, and banking [1]–[4]. Random number generators (RNGs) are categorized in three major types: Pseudo-RNG (PRNG), Chaotic-RNG (CRNG) and True-RNG (TRNG). Pseudorandom numbers are generated by employing mathematical algorithms with an initial state called seed. A PRNG provides deterministic, periodic and predictable sequences; therefore, it cannot be truly random. However, while algorithmically generated pseudorandom numbers can be used for some applications but due to the predictable structure of them it cannot be used in cryptography, statistical calculations, and simulations. CRNGs are based on chaotic physical processes. Chaos is the long-term non-predictive behavior observed in certain nonlinear dynamical systems due to the sensitivity of the output trajectory to the initial conditions. Although chaotic systems are deterministic and can be described using simple differential equations, their

output trajectory can only be determined given the exact initial state of the system. As in practice, the initial state of the system is only known with uncertainty due to the random environmental noise. The perturbations grow exponentially, leading to unpredictability [5]. For instance, CRNGs exploit thermal noise [6], optical fluctuations of laser radiation [7], and jitter of oscillators in integrated circuits [8]. A TRNG is based on a physical random process rather than computational algorithms; however, only a physical way does not sufficiently guarantee the true randomness, it is widely accepted that the core of any TRNG must be an intrinsically non-deterministic physical process. Current theories imply that the only way to realize a TRNG which can be scientifically proven to be non-deterministic is to use the intrinsic randomness of quantum decisions, since the occurrence of each possible result is unpredictable and unreproducible [9]. Quantum random number generators (QRNGs) are a special class of physical random data sources, whose randomness is established from elementary quantum processes, and quantum optical processes are ideal for realizing such devices because of their relative ease of implementation. The foundation of quantum physics is the unpredictability factor, and therefore, it guarantees the randomness for QRNG. In an ideal QRNG, the probability of "1" and "0" is equal, i.e., $P(0) = P(1) = 0.5$, and generated bits are statistically independent; therefore, the autocorrelation coefficients between them are zero. However, all physical random number generators show some deviation from the mathematical ideal of statistically independent and uniformly distributed bits. Therefore, the raw output of QRNGs tends to be slightly biased and may even deteriorate over time. More recently, a variety of QRNGs was developed using different types of quantum randomness. They all exhibit specific advantages, but often, they also have one or more disadvantages like low data rates, poor quality of raw random numbers either due to the bias or correlations along the bit sequence, and/or complex implementations [10]. Photon detection time [11]–[15], quantum phase fluctuations [16]–[18], detection of direction or polarization of a single photon [19]–[22], shot noise in vacuum states [23], photon number decision of weak laser pulses [24], photon detection by array of the photodetectors [25], and radioactive decay [26] are some important non-deterministic physical resources and approaches for QRNGs. Besides the speed and quality of the generated random bits, system complexity, cost, reliability, and sensitivity to control parameters are also other important factors for a successful QRNG. Since no practical true single-photon source exists at present, QRNGs based on single photons usually use a weak discrete laser or LED source to approximate a single photon source. However, a laser-based setup usually is complex, expensive, large in size, and not implementable as a monolithic integrated circuit.

2. State of the Art

Recently, the measurement of photon detection time is used as a method for generating high-speed random numbers. In this type of QRNG, photons emitted from a continuous-wave laser diode or an LED is detected by a single-photon detector, and the time intervals between successive detection events are recorded as the raw data [11]–[15]. All these setups are LED-based and employ a non-silicon LED as discrete component for generation of single photons, and therefore they cannot be implemented in conventional CMOS technology so that it makes the setup complex and large in volume. In this paper, we present and demonstrate a novel monolithic optoelectronic setup to generate quantum random bits, which utilizes the method of measuring single-photon detection time. In our design, a Si-CMOS-LED with a single-photon avalanche photodiode (SPAD) integrated in standard 0.35 μm high-voltage (HV) CMOS technology. The Si-LED operates in the reverse avalanche mode and emits photons for detection towards the SPAD. Also, all our setup has the potential of being fully integratable with standard CMOS integrated circuitry with no adaptation of the CMOS design and processing procedures. It leads to a low-cost, simple and more compact set up. However, similar to many of the other methods, some post-processing of the raw bits was needed to produce random bits of high enough quality. The paper is organized as follows: After an overview to the physical mechanisms underlying on SPADs and Si-LED, the experimental setup of the proposed QRNG is presented.

Then, an effective post-processing algorithm for applying on raw bits is given. Subsequently, experimental results and some important statistical analyses on generated random number streams are expressed. Finally, concluding remarks are given.

3. Physical Mechanisms in SPADs and Si-LEDs

The QRNG acts based on statistical detection of the quantum events by means of a highly sensitive quantum detector, namely an SPAD. A single-photon avalanche diode is implemented as a p-n junction reverse biased well above its breakdown voltage so that it operates in Geiger mode [27]. At this bias, the electric field in the depletion layer is so high that a single charge carrier injected into the active region can trigger a self-sustaining avalanche, thus current rises swiftly and creates a macroscopic current pulse in the range of a few milliamperes through the device. The current continues until the avalanche is quenched by lowering the bias voltage down to the breakdown voltage, the lower electric field is no longer able to accelerate carriers to impact-ionize; therefore, the current ceases. Triggering and creating an avalanche in an SPAD is a statistical process [28]. Thermally and tunneling generated carriers within depletion layer can also trigger avalanche pulses that are indistinguishable from actual radiation-triggered pulses. The frequency of generation of these spurious pulses is known as dark count rate. Moreover, some charge carriers that were trapped by crystal defects are released at a certain time and re-trigger the SPAD. This afterpulsing has a probability which is a function of trap density, the number of carriers generated during an avalanche and of the release time of these carriers [29]. Afterpulsing can inhibit high-frequency operation of SPADs. Both dark count and afterpulsing degrade the performance of SPADs and create some deviations and bias in generated random bits.

We utilized a Si-LED as photon generation source in our design. Light emission from silicon devices has been realized in reverse-biased p-n avalanche structures [30]. Si-LEDs emit light and generate photons in the visible and near-infrared spectrum. Various theories have been put forward in order to explain the phenomenon. These include phonon-assisted intraband relaxation phenomena, as well as phonon interband recombination processes [31]. In-depth theoretical modeling and experimental evidence indicated that the dominant photonic generation processes may be from intraband phonon-assisted relaxation processes, mainly in the conduction band [32]. The intensity of emitted light is a function of the product of the probability of recombination and the distribution functions for electrons and holes, considering acoustical, optical and ionizing scattering. This optical radiation comes from the kinetic energy loss of carriers generated by impact ionization colliding with immobile charge centers in the avalanche region [33]. Si is an indirect semiconductor and suffers from low quantum efficiency in generation of the light; thus, the output optical power is weak.

Concerning an operating temperature up to e.g., 75 °C, it has to be noted that both Si-LED and SPAD are based on the avalanche effect, from which it is well known that the breakdown voltage increases with temperature. For constant bias voltages, it therefore has to be expected that the pulse rate of the random pulse generator (RPG) will decrease with increasing temperature. As a consequence, the bias voltages of Si-LED and SPAD have to be increased with temperature to keep the pulse rate constant. In practice, a control circuit will have to be implemented to guarantee the correct operation of the RPG in a wide operating temperature range.

4. Experimental Setup

The block diagram of the proposed setup for QRNG is shown in Fig. 1. The main section of this setup is the random pulse generator (RPG) consisting of Si-LED and ring SPAD which generate the random pulses. The space between Si-LED and SPAD has a width of about 10 μm. To be able to avoid reflections in the cables and to perform measurements at the output of the RPG an impedance matching to 50 Ohm was implemented. A National Instruments myRIO module with an XILINX ZYNQ-7010 field-programmable gate array (FPGA) containing a real-time processor was connected. This processor used a high speed pulse generator to generator its 40-MHz clock. A phase-locked loop (PLL) was implemented to generate a synchronous 250-MHz clock.

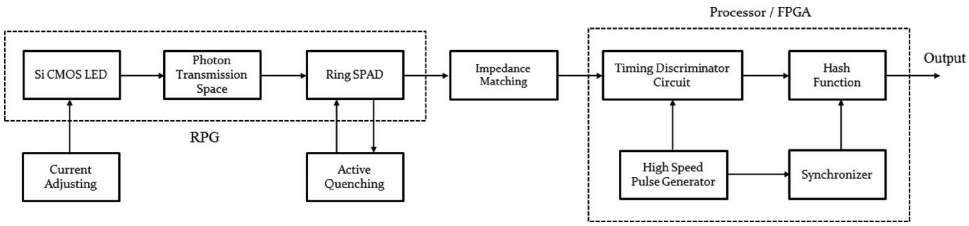


Fig. 1. Block diagram of the proposed setup for QRNG.

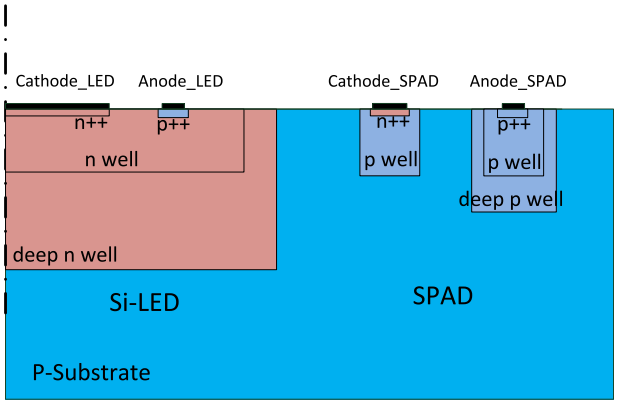


Fig. 2. Cross section of the random pulses generator (RPG).

Fig. 2 shows the cross section of the random pulses generator (RPG). As shown in this figure, the Si-LED as photon generation source located in the center with circular shape and the SPAD as highly sensitive photodetector is around it in a ring structure. The Si-LED consists of eight equidistant p^{++} regions on a circle with a radius of 12.35 μm. The multiplication zone of the SPAD is formed at the n^{++}/p-well junction. The radius of the n^{++} cathode ring of the SPAD is 22.9 μm. An n^{++} width of 1.2 μm was layouted. The light is emitted at the p^{++}/n-well junction. The RPG was fabricated in 0.35 μm HV-CMOS technology. Breakdown voltages of Si-LED and SPAD are 9.7 V and 10.9 V, respectively. The Si-LED in reverse mode generates the photons based upon the avalanche mechanism, so that the photon flux can be controlled with the bias current of the LED. The quantum efficiency of the Si-LED is very low and, therefore, the generated photon flux is weak. We used the shallow junctions with p^{++} doping and n-well to increase the electric field strength and consequently increment the quantum efficiency of the device. In addition, a deep n-well is placed around the n-well in order to isolate the LED better from the substrate and from the SPAD. The obtained results show that some photons reach the depletion region of the SPAD and trigger it. The SPAD operates in Geiger mode which means that, once triggered the avalanche current keeps on flowing, thus rendering the device useless for subsequent detections. A fast quenching circuit is necessary, which can quickly sense the leading edge of the avalanche current of the SPAD, generate a standard output pulse, immediately reduce the SPADs bias voltage, and then bring the SPAD back to its original quiescent state [34]. We utilized the active quenching approach for the SPAD in our setup. The active quencher delivered an adjustable excess bias voltage to the SPAD. The active quenching circuit allowed to vary the deadtime of the SPAD for durations larger than 30 ns. This circuit was fabricated in 0.35-μm HV-CMOS technology. All experiments were performed at room temperature (25 °C). As shown in Fig. 3, the output response of the active quenching circuit is a stream of random pulses. Time intervals between two sequential pulses detected by the SPAD are independent random variables. The photon statistics for an attenuated LED light source is approximately Poissonian. It can be shown that Poissonian events are occurring independently of each other

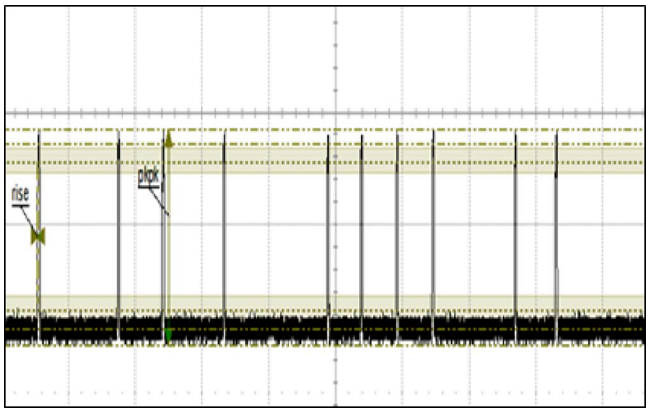


Fig. 3. Generated random pulses in the output of active quenching circuit, displayed on an oscilloscope at 500 ns/div and 1 V/div.

and that the waiting-time distribution takes the shape of a decaying exponential. Therefore, a time-wise random stream of the electrical pulses is obtained from the random pulse generator (RPG) which generates electrical pulses whose waiting-times obey an exponential distribution function. The basic idea of the method of extracting random bits is to consider a pair of non-overlapping random time intervals (t_{2i-1}, t_{2i}) which are defined with sequential random events. To digitize the random time intervals (t_{2i-1}, t_{2i}), we employed a XILINX ZYNQ-7010 FPGA. The digital outputs are defined as follows:

$$\text{Output} = \begin{cases} 1, & \text{if} \quad t_{2i-1} > t_{2i} \\ 0, & \text{if} \quad t_{2i-1} < t_{2i} \\ \text{Abandoned}, & \text{if} \quad t_{2i-1} = t_{2i} \end{cases} \tag{1}$$

$$i = 1, 2, 3, \ldots$$

and we have

$$P(t_{2i-1} > t_{2i}) = P(1) = 0.5 - \varepsilon \tag{2}$$

$$P(t_{2i-1} < t_{2i}) = P(0) = 0.5 + \varepsilon \tag{3}$$

where ε is the bias of the bits in raw data. The rate of generated pulses at the output of the active quenching circuit is dependent on the bias current of the LED. To increase the speed of QRNG it is necessary to reduce the mean value of time intervals between sequential pulses by increasing the bias current of the Si-LED. The measurement results on photocurrent of the avalanche photodiode (APD) in linear mode as an optical sensor shows that at lower reverse currents, the relationship of the generated photocurrent versus LED current is approximately linear, but as the LED current increases, the photocurrent shows a trend of saturation; therefore, this mechanism in Si-LED limits the speed. On the other hand, maximum speed of the generated random bits is bounded by the deadtime of the SPAD and also FPGA response, too. With an LED current of 5 mA and excess bias voltage of 2.5 V, a raw random stream with bit rate of 2 Mbit/s is obtained for described RPG and the used setup.

5. Post-Processing

Statistical independence and uniform distribution of generated random bits are two fundamental factors in an ideal RNG. However, physical RNG might suffer from correlations for short lags and bias due to the random bit extraction process, as well as limitations of the used instrumentation [35]. In other words, in practice, hardware-based random number generators sometimes

have some deflection from ideal state and, therefore, in raw bit streams, are created bias, which is a deviation of the 1-probability from 1/2. With appropriate modeling, one can prove the generated bits to be information-theoretically random after certain data post-processing called randomness extraction. Post-processing on generated random bits can remove or at least reduce bias effects. Common techniques used in the post-processing step include hashing or blockwise XOR-ing. A popular method is to use a linear feedback shift register (LFSR) where the bits from the true random number generator are XORed to the feedback value computed according to the feedback polynomial. This method tries to make use of the good statistical properties of linear feedback shift registers. When the bits from the physical source are nonconstant but biased, some of the bias is removed, but additional dependencies between output bits are introduced by LFSR method [36]. However, in this paper we utilize a linear code extractor with a fixed number of inputs and capability of more simple hardware implementation which can significantly improve the quality of the generated random bits. Compared to [14] and [15], where the Von Neumann corrector was applied, the XOR hashing suggested here results in a higher throughput.

For $X = [x_0, x_1, \ldots, x_{n-1}]$ and $Y = [y_0, y_1, \ldots, y_{m-1}]$, any linear binary corrector mapping n bits to m bits is defined as the product of the vector X by the binary matrix $G = [g_{i,j}]$ as $GX = Y$. The matrix G is indicated as generator matrix of a $[n, m, d]$ linear code. Where d is the minimal distance of the code and any nonzero linear combination of output bits is the sum of at least d input bits. Any linear $[n, m, d]$-code provides a linear corrector with compression ratio of n/m and an estimation of the upper bound of the output bias. The upper bound of this bias is $2^{d-1}\varepsilon^d$. To reduce the output bias, a linear corrector with large enough d is suitable. The hardware implementation of the linear corrector is efficiently achieved as simple circuit using XOR gates and a shift register. There are no linear binary codes of length 16 and dimension 8 with minimal distance greater than 5. In these conditions, to minimize output bias, we must use nonlinear correctors and therefore, a complex hardware implementation will be needed. According to the above mentioned, to obtain a linear corrector with minimum output bias and a simple hardware structure, we considered the [5], [8], [16] linear code for our post-processing, with compression factor of 2 and output bias less than $16\varepsilon^5$.

Let $X = [x_0, x_1, \ldots, x_{15}]$ be the input bits and $Y = [y_0, y_1, \ldots, y_8]$ be the output bits for post-processing. We split the X vector into the two bytes $X = [A|B]$, and the output vector Y is written as follows [36]:

$$Y = A \oplus RL(A, 1) \oplus RL(A, 2) \oplus RL(A, 4) \oplus B \tag{4}$$

where $RL(A, k)$ denotes the rotation in A to the left by the number of k bits and is the XOR operator in the digital domain. We can convert above equation to the eight expressions for the output bit stream as follows:

$$\begin{aligned} y_7 &= x_{15} \oplus x_3 \oplus x_5 \oplus x_6 \oplus x_7 \\ y_6 &= x_{14} \oplus x_2 \oplus x_4 \oplus x_5 \oplus x_6 \\ y_5 &= x_{13} \oplus x_1 \oplus x_3 \oplus x_4 \oplus x_5 \\ y_4 &= x_{12} \oplus x_0 \oplus x_2 \oplus x_3 \oplus x_4 \\ y_3 &= x_{11} \oplus x_7 \oplus x_1 \oplus x_2 \oplus x_3. \end{aligned} \tag{5}$$

These expressions were implemented in the FPGA as hash-function for whiting the input bit streams. To generate a file with 1 Gbit random bits took 1000 seconds with the LED current of 5 mA. This file size of 1 Gbit was chosen because this is the standard file size of the NIST test suite. With a $[n, m, d]$ linear code and input bias ε, the minimal entropy $H_{\min}$ of the output bit streams versus the probability of a "1" is given by [37]:

$$H_{\min} = m - \mathrm{Log}_2(1 + 2^{m+d}\varepsilon^d). \tag{6}$$

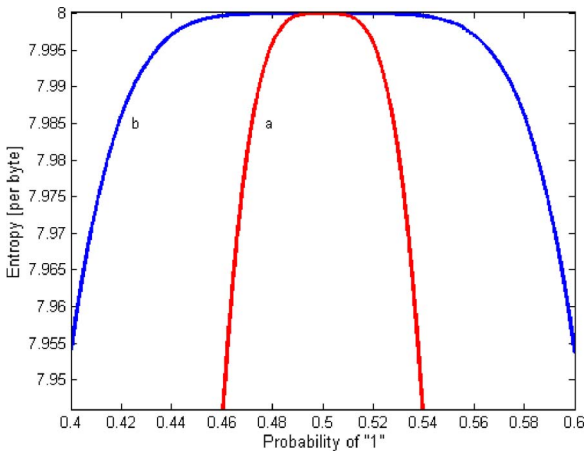


Fig. 4. Min-entropy of the output bit streams versus the probability of "1"-bits. (a) Before post-processing. (b) After post-processing.

For [5], [8], and [16] in our post-processing, the minimal entropy of the output bit streams is drawn in Fig. 4. As can be seen, the proposed algorithm for post-processing modifies the output entropy and reduces the existing bias in the raw data significantly. In the next section, we compute some important statistical properties for generated random streams before and after post-processing, and subsequently obtained results are compared.

6. Experimental Results

6.1. Characterization

An SPAD has several sources of non-idealities such as deadtime, dark counts, and afterpulsing. Deadtime relates to the time required to return to the initial state after an avalanche has occurred; it is the minimum time between detection pulses. In other words, detector deadtime is caused by a minimum recovery time required for the detector being able to detect the next photon after a previous detection event. Usually, detector deadtimes are in the range of several tens to hundreds of nanoseconds. In our setup, the active quenching circuit delivers the minimum deadtime of 30 ns for the SPAD. As explained before, dark counts are random events due to two quantum mechanisms, where thermally generated carriers or carriers generated by band-to-band tunneling trigger the avalanche. This is generally a Poissonian process and it can be characterized through the dark count rate (DCR), a mean or median event rate, which is a function of excess bias voltage, temperature, and the active area of the SPAD. During the measurement of dark count rate the SPAD is kept in complete darkness. However, the case of simple thermal generation as intrinsic noise of the SPAD should be independent of hold-off duration. Instead, the increase of the dark count rate at short hold-off durations predicates a clear sign of afterpulsing [28]. To measure the thermally dark count rate it is necessary to set conditions where the effect of afterpulsing is negligible. The simplest way to obtain the thermally dark count rate is measurement in a condition where the deadtime is extended by enforcing a suitable hold-off time, long enough to allow all the trapped carriers to be released [38]. Fig. 5 shows the measured dark count rate as a function of excess bias voltage for a hold-off time of 30 ns and at a temperature of 298 K. As can be seen, for the excess bias voltage of 2.5 V used in our measurement the DCR is less than 15 kHz.

Afterpulsing is characterized by the afterpulsing probability, a parameter that relates the probability of secondary and higher-order avalanches to the excess bias voltage and hold-off time. The origin of afterpulsing and its characteristics depend on the detector conditions and temperature. The total afterpulsing probability as a function of hold-off time measured at an excess bias

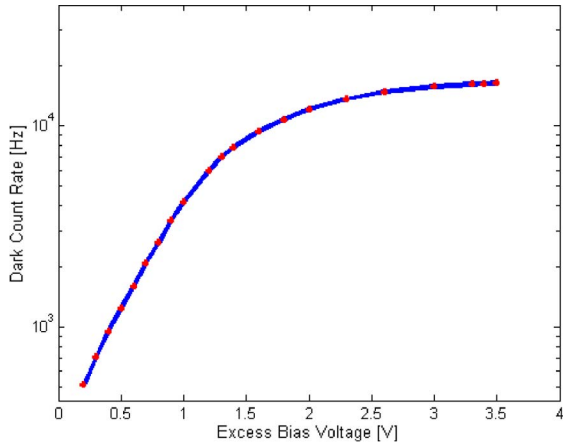


Fig. 5. Dark count rate as a function of excess bias voltage at 298 K.

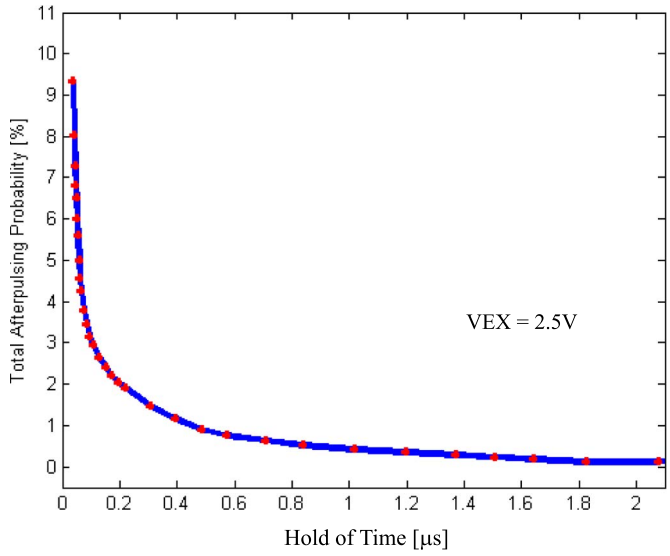


Fig. 6. Total afterpulsing probability versus the hold-off time at 298 K.

voltage of 2.5 V and a temperature of 298 K is given in Fig. 6. In these measurements, the deadtime may be varied from a minimum of 30 ns to a few microseconds. As can be seen, the afterpulsing probability declines for longer separation times between the avalanches. This phenomenon is generally attributed to the exponential decay of trapped carriers from trap states in the avalanche region. The obtained results show that the afterpulses are substantially reduced after 200 ns, but it takes approximately 2 μs to reach the low intrinsic afterpulse-free DCR level pertaining to that bias voltage. The photon count rate in the RPG is 4 MHz that corresponds to an average interval between two counts of 250 ns, which is much longer than the minimum deadtime of 30 ns of the SPAD resulting from the active quenching circuit used. Therefore, the quenching circuit did not limit the photon counting rate, and for the deadtime of 30 ns the probability of afterpulsing was less than 9%. Although afterpulses are correlated with genuine photon-detection events, they can be negligible when bit extraction rate is high, compared to the correlation time, whereas the correlation time characteristics between photon detections and subsequent afterpulses is on the order of a few microseconds; therefore, the probability of counting an afterpulse after some photon detection hardly changes, leading to a nearly uniform

TABLE 1
Results of the ENT Tests Suite for Before and After Post-Processing

Before post-processing
Entropy = 7.999506 bits per byte Arithmetic mean value of data = 0.502164 (random = 0.5000000) Monte Carlo value for π is 3.143791768 (error 0.07 percent) Serial Correlation coefficient = - 0.001815 (totally uncorrelated = 0.000)
After post-processing
Entropy = 8.000000 bits per byte Arithmetic mean value of data = 0.5000025 (random = 0.5000000) Monte Carlo value for π is 3.141497690 (error 0.003 percent) Serial Correlation coefficient = - 0.000057 (totally uncorrelated = 0.000)

distribution of afterpulses on the photon-detection time scale. However, bits produced by our setup at 1 Mbit/s show no measurable correlations.

6.2. Statistical Properties

Some important statistical properties of generated random bit streams before and after applying the proposed post-processing algorithm are computed using ENT [39]. ENT is a series of basic statistical tests which evaluate the random sequence in some elementary features such as the equal probabilities of ones and zeros, the serial correlation, and Monte Carlo simulation for estimation of π. The testing results with ENT are presented in Table 1. In this computation, the input binary stream to the extractor is 2 Gbit so that after post-processing it is reduced to 1 Gbit. From the results, it can be seen that before post-processing the initial entropy is 7.999506 bits per byte, the numerical error in Monte Carlo simulation for calculation of π is less than 0.07 percent and the probability of ones or zeros is 0.502164. However, after applying the proposed post-processing algorithm our QRNG indicates the full entropy of 8.000000 bits per byte and it can generate ones and zeros with a probability very close to 1/2. Moreover, the serial auto-correlation coefficient in the order of 10^{-5} and the numerical error for estimation of π less than 0.003 percent are achieved. Therefore, the post-processing algorithm applied on raw data improves the quality of the quantum random numbers generated in our setup considerably.

6.3. Randomness Tests

To evaluate the quality of the generated random binary streams, we performed all statistical tests of randomness from DIEHARD [40] and STS [41] on numerous bit streams of 1 Gbit each on a personal computer. DIEHARD is widely considered as one of the best strengthened randomness testing battery because it is most sensitive to various problems possible in RNG. It consists of 15 independent tests with an outcome of one or more p-values. A sequence can successfully pass the test if the p-value is larger than 0.01 and less than 0.99 ($0.01 < p\text{-value} < 0.99$). Testing results on a typical sequence of 1 Gbit with DIEHARD are shown in Table 2. For the cases of multiple p-values, a Kolmogorov–Smirnov (KS) test is used to obtain a final p-value, which measures the uniformity of the multiple p-values. STS from the National Institute of Standards and Technology (NIST) is another powerful randomness test with 14 independent statistical tests. Obtained results of the randomness testing with STS on binary streams are illustrated in Table 3. As is shown in Tables 2 and 3, all randomness tests were successfully passed for both DIEHARD and STS suites. It is worth mentioning to say that we tested a lot of bit streams as long as 1 Gbit with the two statistical tests suites which were passed all randomness tests.

TABLE 2

Results of the Diehard Statistical Tests on 10^9 bit Pattern of Binary Bits From the QRNG

Test	p-value	Result
Birthday Spacing	0.236351 [KS]	SUCCESS
Overlapping Permutations	0.556270	SUCCESS
Ranks of 31×31 Matrices	0.736638	SUCCESS
Ranks of 32×32 Matrices	0.785979	SUCCESS
Ranks of 6×8 Matrices	0.583348 [KS]	SUCCESS
Monkey Tests on 20-bit Words	0.445804 [KS]	SUCCESS
Test OPSO	0.775634 [KS]	SUCCESS
Test OQSO	0.756108 [KS]	SUCCESS
Test DNA	0.557120 [KS]	SUCCESS
Count 1's in Stream of Bytes	0.693431	SUCCESS
Count 1's in Specific of Bytes	0.371287 [KS]	SUCCESS
Parking Lot Test	0.508434 [KS]	SUCCESS
Minimum Distance Test	0.972883 [KS]	SUCCESS
Random Spheres Test	0.951424 [KS]	SUCCESS
Squeeze Test	0.369728 [KS]	SUCCESS
Overlapping Sums Test	0.293293 [KS]	SUCCESS
Runs Test (up)	0.334992	SUCCESS
Runs Test (down)	0.397404	SUCCESS
Craps Test No. of Wins	0.683578	SUCCESS
Craps Test throws / game	0.373586	SUCCESS

TABLE 3

Typical Results of the NIST Statistical Test Suite on a Sequence of 1 Gbit

Test	p-value	Result
Frequency	0.299251	SUCCESS
Block Frequency	0.213309	SUCCESS
Cumulative Sums	0.350485	SUCCESS
Runs	0.407091	SUCCESS
Longest Run	0.253551	SUCCESS
Rank	0.468595	SUCCESS
FFT	0.407091	SUCCESS
Universal	0.739918	SUCCESS
Serial	0.671779	SUCCESS
Linear Complexity	0.178278	SUCCESS
Non Overlapping Template	0.804897	SUCCESS
Approximate Entropy	0.691881	SUCCESS
Random Excursions	0.732648	SUCCESS
Random Excursions Variant	0.596399	SUCCESS

6.4. Statistical Analysis

In addition to the randomness tests, we estimated the bias and serial correlation of the output bit streams. The probability of "1" was computed from 100 bit streams of 1 Gbit each in order to investigate the bias existing in the generated random bits. The distribution of bias versus the bit streams is shown in Fig. 7. Also, results of the statistical analyses on distribution of bias are shown as a histogram in Fig. 8. As predicted according to the central limit theorem, the final distribution of bias is a Gaussian function with the peak value around zero and the standard deviation of $\sigma_{\text{mean}} = 0.5/\sqrt{N} = 1.58 \times 10^{-5}$ for $N = 1 \times 10^9$ bits. The serial autocorrelation

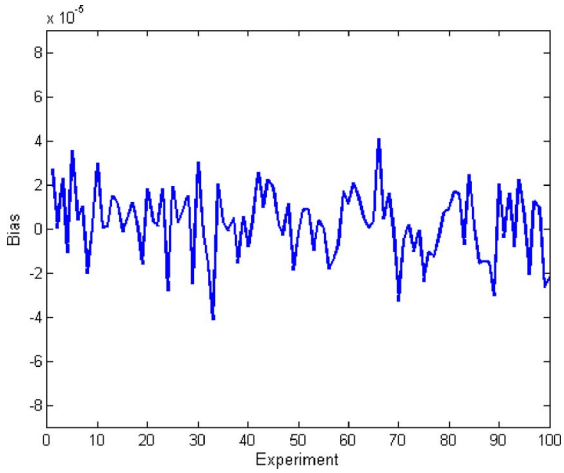


Fig. 7. Measured bias for 100 binary streams of 10^9 bits each. The mean value is close to zero, and the standard deviation is less than 2×10^{-5}.

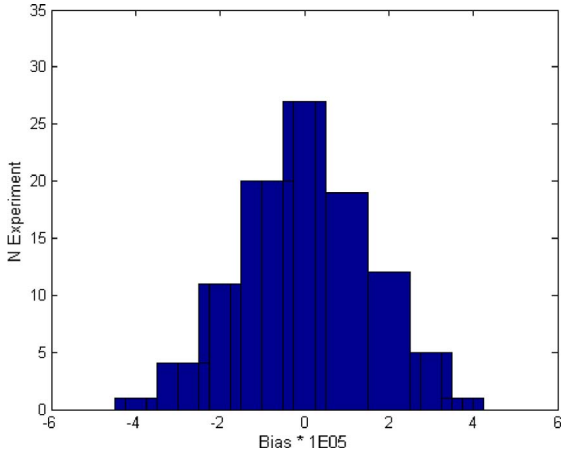


Fig. 8. Results of the statistical analyses on bias distribution of the bits as a histogram.

coefficients for the bit stream with lag $k \geq 1$ were also calculated according to the following equation [42]:

$$\rho_k = \frac{1}{M-k} \times \frac{\sum_{i=1}^{M} b_i b_{i+k} - \mu^2}{\sigma^2} \tag{7}$$

where b_i is the ith bit of the bit stream, and μ and σ^2 are the mean value and variance of the random b_i, respectively. The first 1200 coefficients of the serial autocorrelation for a typical sequence of 1 Gbit is shown in Fig. 9. Autocorrelation coefficients are well distributed around zero with a standard deviation of 5.94×10^{-5}. In this figure, the horizontal lines define the statistical levels of $\pm 3\sigma$. Calculated coefficients show no correlation in the data streams.

6.5. Stability

To accumulate 100 bit streams each with a size of 1 Gbit took almost 30 hours. Within this time period, no degradation was observed. Fig. 7 shows a constant low bias for all 100 experiments with 1 Gbit of data each. The pulse rate of the random pulse generator was still the same after 30 hours.

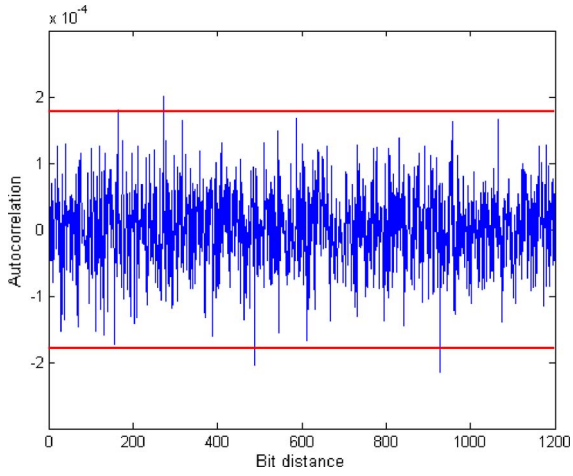


Fig. 9. Serial autocorrelation coefficients for a typical sequence of 10^9 bits with the lag k ranging from 1 to 1200. The horizontal lines indicate the statistical levels of $\pm 3\sigma$.

7. Conclusion

A new monolithically integrated approach to realize a non-deterministic RNG based on quantum effects in photonic emission and detection is presented. A Si-LED as photon generation source located in the center and a ring SPAD around it as a highly sensitive photodetector with active quenching circuit on the same chip can be a suitable structure for generation of the quantum random numbers. This configuration was fabricated in 0.35 μm HV-CMOS technology. The density of the generated random pulses at the output of the active quenching circuit is increased with incrementing the bias current of the Si-LED. A higher generation rate is attainable with further bias current in the Si-LED, lower deadtime in the SPAD, and faster data acquisition hardware. For an LED current of 5 mA the final random bit generation rate up to 1 Mbit/s is achieved. The randomness of our TRNG is physically guaranteed by intrinsic random nature of avalanche phenomena in photon generation in the Si-LED. Our QRNG successfully passed all randomness tests from ENT, DIEHARD, and STS. Moreover, the true randomness is verified by estimation of both correlation coefficients and statistical bias for 100 random bit sequences of 1 Gbit. A simple but highly qualitative post-processing was done in an FPGA. However, this XOR post-processing can easily be implemented in a low-area digital circuit on the same chip, together with the RPG and active quencher resulting in a very compact and cheap QRNG, which is also embeddable as a module in a system-on-chip, which needs reliable, secret, and secure random numbers.

Acknowledgments

The authors would like to thank Dr. H. Mahmoudi for helpful discussions.

References

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*. New York, NY, USA: Springer-Verlag, 2007.
[2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 145–195, Jan.–Mar. 2002.
[3] N. Ferguson, B. Schneier, and T. Kohno, *Cryptography Engineering: Design Principles and Practical Applications*. Indianapolis, IN, USA: Wiley, 2010.
[4] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*. Hoboken, NJ, USA: Wiley, 2008.
[5] A. Beirami and H. Nejati, "A framework for investigating the performance of chaotic-map truly random number generators," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 60, no. 7, pp. 446–450, Jul. 2013.
[6] Y. Yamanashi and N. Yoshikawa, "Superconductive random number generator using thermal noises in SFQ circuits," *IEEE Trans. Appl. Supercond.*, vol. 19, no. 3, pp. 630–633, Jun. 2009.
[7] C. R. S. Williams, J. C. Salevan, X. Li, R. Roy, and T. E. Murphy, "Fast physical random number generator using amplified spontaneous emission," *Opt. Exp.*, vol. 18, no. 23, pp. 23584–23597, Nov. 2010.

[8] B. Sunar, W. J. Martin, and D. R. Stinson, "A provably secure true random number generator with built-in tolerance to active attacks," *IEEE Trans. Comput.*, vol. 56, no. 1, pp. 109–119, Jan. 2007.
[9] A. Alkassar, T. Nicolay, and M. Rohe, "Obtaining true random binary numbers from a weak radioactive source," in *Computational Science and Its Applications*. Berlin, Germany: Springer-Verlag, 2005.
[10] M. Fürst *et al.,* "High speed optical quantum random number generation," *Opt. Exp.*, vol. 18, no. 12, pp. 13029–13037, Jun. 2010.
[11] M. Stipcevic and B. M. Rogina, "Quantum random number generator based on photonic emission in semiconductors," *Rev. Sci. Instrum.*, vol. 78, no. 4, Apr. 2007, Art. ID. 045104.
[12] M. Wahl *et al.,* "An ultrafast quantum random number generator with provably bounded output bias based on photon arrival time measurements," *Appl. Phys. Lett.*, vol. 98, no. 17, Apr. 2011, Art. ID. 171105.
[13] Y. Nie *et al.,* "Practical and fast quantum random number generation based on photon arrival time relative to external reference," *Appl. Phys. Lett.*, vol. 104, no. 9, Feb. 2014, Art. ID. 051110.
[14] S. Burri *et al.,* "Architecture and applications of a high resolution gated SPAD image sensor," *Opt. Exp.*, vol. 22, no. 14, pp. 17573–17589, Jul. 2014.
[15] S. Burri *et al.,* "SPADs for quantum random number generators and beyond," in *Proc. 19th IEEE ASP-DAC*, 2014, pp. 788–794.
[16] H. Guo, W. Tang, Y. Liu, and W. Wei, "Truly random number generation based on measurement of phase noise of a laser," *Phys. Rev. E, Stat., Nonlinear, Soft Matter Phys.*, vol. 81, no. 5, May 2010, Art. ID. 051137.
[17] M. Jofre *et al.,* "True random numbers from amplified quantum vacuum," *Opt. Exp.*, vol. 19, no. 21, pp. 20665–20672, Oct. 2011.
[18] F. Xu *et al.,* "Ultrafast quantum random number generation based on quantum phase fluctuations," *Opt. Exp.*, vol. 20, no. 11, pp. 12366–12377, May 2012.
[19] J. G. Rarity, P. C. M. Owens, and P. R. Tapster, "Quantum random number generation and key sharing," *J. Mod. Opt.*, vol. 41, no. 12, pp. 2435–2444, Dec. 1994.
[20] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, "Optical quantum random number generator," *J. Mod. Opt.*, vol. 47, no. 4, pp. 595–598, Mar. 2000.
[21] *Random Number Generation Using Quantum Physics*, ID Quantique White Paper, 2010. [Online]. Available: http://www.idquantique.com/
[22] T. Durt *et al.,* "Fast quantum-optical random-number generators," *Phys. Rev. A, At., Mol., Opt. Phys.*, vol. 87, no. 2, Feb. 2013, Art. ID. 022339.
[23] A. Marandi, N. C. Leindecker, K. L. Vodopyanov, and R. L. Byer, "All-optical quantum random bit generation from intrinsically binary phase of parametric oscillators," *Opt. Exp.*, vol. 20, no. 17, pp. 19322–19330, Aug. 2012.
[24] M. Ren *et al.,* "Quantum random-number generator based on a photon-number-resolving detector," *Phys. Rev. A, At., Mol., Opt. Phys.*, vol. 83, no. 7, Jul. 2011, Art. ID. 023820.
[25] G. Ribordy and O. Guinnard, "*Method and apparatus for generating true random numbers by way of a quantum optics process*," U.S. Patent 007519641B2, Jun. 7, 2009.
[26] Y. Yoshizawa *et al.,* "Physical random numbers generated by radioactivity," *J. Jpn. Soc. Comput. Stat.*, vol. 12, no. 1, pp. 67–81, Dec. 1999.
[27] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in quenching circuits for single photon avalanche diodes," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 6, pp. 3815–3826, Dec. 2010.
[28] S. Tisa, F. Guerrieri, and F. Zappa, "Variable-load quenching circuit for single-photon avalanche diodes," *Opt. Exp.*, vol. 16, no. 3, pp. 2232–2244, Feb. 2008.
[29] E. Vilella and A. Dieguez, "A gated single-photon avalanche diode array fabricated in a conventional CMOS process for triggered systems," *Sens. Actuators A, Phys.*, vol. 186, pp. 163–168, Oct. 2012.
[30] A. L. Lacaita, F. Zappa, S. Bigliasrdi, and M. Manfredi, "On the Bremsstrahlung origin of hot-carrier-induced photons in silicon devices," *IEEE Trans. Electron Devices*, vol. 40, no. 2, pp. 577–582, Mar. 1993.
[31] N. Akil, S. E. Kerns, D. V. Kerns, Jr., A. Hoffmann, and J. P. Charles, "A multi-mechanism model for photon generation by silicon junctions in avalanche breakdown," *IEEE Trans. Electron Devices*, vol. 46, no. 5, pp. 1022–1027, May 1999.
[32] J. Bude, N. Sano, and A. Yoshii, "Hot carrier luminescence in silicon," *Phys. Rev. B, Condens. Matter*, vol. 45, no. 11, pp. 5848–5856, Mar. 1992.
[33] K. Xu and G. P. Li, "A novel way to improve the quantum efficiency of silicon light-emitting diode in a standard silicon complementary metal–oxide–semiconductor technology," *Appl. Phys.*, vol. 113, no. 10, Mar. 2013, Art. ID. 103106.
[34] Z. Lixia *et al.,* "Active quenching circuit for a InGaAs single-photon avalanche diode," *J. Semicond.*, vol. 35, no. 4, pp. 1–6, Apr. 2014.
[35] S. Tisa, F. Villa, A. Giudice, G. Simmerle, and F. Zappa, "High-speed quantum random number generation using CMOS photon counting detectors," *IEEE J. Sel. Topics Quantum Electron.*, vol. 21, no. 3, May/Jun. 2014, Art. ID. 6300107.
[36] M. Dichtl, "Bad and good ways of post-processing biased physical random numbers," in *Fast Software Encryption*, vol. 4593, Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2007, pp. 137–152.
[37] P. Lacharme, "Post-processing functions for a biased physical random number generator," in *Fast Software Encryption*, vol. 5086, Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 2008, pp. 334–342.
[38] A. C. Giudice, M. Ghioni, S. Cova, and F. Zappa, "A process and deep level evaluation tool: Afterpulsing in avalanche junctions," in *Proc. IEEE ESSDERC*, Estoril, Portugal, 2003, pp. 347–350.
[39] J. Walker. [Online]. Available: http://www.fourmilab.ch/random/
[40] G. Marsaglia, *The Marsaglia random number CDROM including the diehard battery of tests of randomness*, DIEHARD, 1995. [Online]. Available: http://www.stat.fsu.edu/pub/diehard/
[41] *For the NIST Statistical Tests Suite*. [Online]. Available: http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html
[42] D. E. Knuth, *The Art of Computer Programming: Semi-Numerical Algorithms*, 3rd ed., vol. 2. Boston, MA, USA: Addison-Wesley, 1997.