Quantum Random-Number Generator Based on Tunneling Effects in a Si Diode

Haihan Zhou,[1,†] Junlin Li,[2,†] Weixing Zhang,[2] and Gui-Lu Long[2,*]

[1] *Department of Physics, North Carolina State University, Raleigh, North Carolina, USA*

[2] *Physics Department, Tsinghua University, Beijing, People's Republic of China*

(Received 27 May 2018; revised manuscript received 29 December 2018; published 25 March 2019)

Previously, we developed a set of photon-free quantum random-number generators (QRNGs) using In-Ga-As single-photon avalanche diodes. We exploited the stochastic property of the quantum tunneling effect. Here, we utilize tunneling signals in Si diodes to generate random numbers. In our experiment, instead of applying periodic pulses to diodes, we apply a fixed voltage and detect the random time intervals between adjacent tunneling signals. This Si QRNG generates raw data with its ratio of min-entropy over ideal entropy reaching 0.98, which is higher than many other QRNG configurations. The final data rate is 6.98 MB/s and could reach 23 MB/s with a more efficient cooling system.

DOI: 10.1103/PhysRevApplied.11.034060

I. INTRODUCTION

Random sequences are widely applied in physics, mathematics, computer science, and many other subjects. In many fields like machine learning [1], cryptography [2], quantum computation [3], and quantum information [4], true randomness is indispensable. There have been many studies on random-number generation [5] and randomness examination [6].

Quantum random-number generator (QRNG) is also an appealing field. Schemes of QRNG usually take advantage of uncertainty and randomness, coming from basic principles in quantum mechanics [7], to develop eligible random-number generation systems. In 2000, the "path choice" of a single polarized photon passing a polarized beam splitter was applied to establish a quantum random-number generator [8]. Later, various schemes were designed and different systems were utilized as random sources, for example, the arrival time of random photons [9,10], phase fluctuation of the vacuum state [11], and quantum phase noise [12]. The generation rates of quantum random numbers in different schemes vary too. Most generators based on discrete signals cannot reach a rate higher than 100 MB/s, while others can reach several GB/s [13] or even more.

In addition to the generation rate, true randomness is another crucial property of these random number generators. Bell's inequality was applied to QRNG designs to ensure true randomness. In 2015, a self-testing quantum random-number generator was first brought up by Lunghi and Bowles. They designed a discrete quantum random-number generator that can continuously measure its own output entropy via estimation of a "dimension witness" [14] and, consecutively, conduct the postprocessing based on the entropy. Following this work, Ma proposed a concept of semi-self-testing QRNG and designed experiments in a single-photon system [15]. In our experiment, we use the dark count signals of silicon diodes as a random source, which is the "noise" in the traditional framework. Previously, some works that utilize noise in superlattice systems as a random source have proven the feasibility of a random-number generator based on noise [16,17].

In our previous study [18], we utilized the randomness of quantum tunneling effects in In-Ga-As diodes as the randomness source. According to collisional ionization theory, we can estimate the tunneling probability in each part of the separated absorption, grading, charge, and multiplication (SAGCM) In-Ga-As/InP single photon avalanche detector (SPAD) [19]. We managed a 15-MB/s quantum random-number generator. In this paper, we still exploit the quantum tunneling effect, but in silicon diodes. Si SPAD has a simpler structure as there is no band-gap difference between the absorption layer and multiplication layer [20]. Furthermore, we simplify the experiment setup by applying a fixed direct current to the silicon diodes and measure the time intervals between adjacent tunneling signals. After data preselection, the min-entropy could reach 9.8 every 10 bits. The output data pass National Institute of Standards and Technology (NIST) [21] and Diehard tests [22] after a simple randomness extraction [23].

In comparison with other QRNGs, our Si-based device has some advantages. Firstly, as mentioned before, it is based on the tunneling effects in Si diodes; thus, no photon source is used and the whole system is quite smaller than other QRNGs. It is simpler, portable, and integratable. For example, many current schemes produce a higher

*gllong@tsinghua.edu.cn

†Junlin Li and Haihan Zhou contributed equally to this work.

generation rate. Yet, they are set up on an optical table. There are many real-life situations in which we do not have much space to place such QRNGs.

Another point is that we have our real-time preselection and postprocessing programs integrated into an field-programmable gated array (FPGA) inside our device. This means that even the dc voltage applied to the Si diodes fluctuates or, more generally, the tunneling probability changes, and our output is stable. Our data are divided into 1024 groups. This segmentation only depends on the average data rate, which can easily be obtained from our FPGA. Thus, we can make real-time adjustments even if the bias voltage fluctuates.

II. COLLISIONAL IONIZATION IN SI SPAD

Properties of free charge carriers in semiconductor diodes have been studied since the 1950s [24–26]. Collisional ionization is one of the most widely accepted theories [27] when modeling these charge carriers. McIntyre proposed this theory in the 1960s. Free carriers like electrons or holes collide with each other while traveling in semiconductors. These collisions excite new carriers.

In order to make these collisions happen, there should be some initialized carriers in diodes. Generally, initiation of collisions can be photon absorption, thermal excitation, tunneling effects, and afterpulse effects in the semiconductor diodes. Here, in our scheme, we suppress other factors to ensure that the tunneling effect is the only source that starts collisional ionization.

Most Si SPADs are *p-i-n* diodes; the absorption layer has the same bandwidth with the multiplication layer. This simple structure makes the analysis of tunneling probability much easier than In-Ga-As/InP SPAD.

According to McIntyre, electrons and holes have two certain collision probabilities per unit length when traveling in semiconductor diodes α_e and α_h. Then, we can calculate the mean collision times $M(x)$ at a specific position inside the intrinsic layer of a *p-i-n* diode:

$$M(x) = \frac{\exp\left(-\int_x^L \alpha_e - \alpha_h dx'\right)}{1 - \int_0^L \alpha_e \exp\left(-\int_{x'}^L \alpha_e - \alpha_h dx''\right) dx'}. \tag{1}$$

Here, 0 and L represent the position of the interface of the intrinsic layer with the p layer and n layer. This equation was first derived by McIntyre [28].

Collision ionization theory points out that the generation of new free carriers inside a diode is probabilistic. This theory also gives us a good estimation of generation probability. Another stochastic process is the avalanche of carriers. The probability of avalanche caused by a single carrier can also be calculated in this model:

$$P_{\text{ava}}(x) = \frac{P_{\text{ava}}(0) f(x)}{P_{\text{ava}}(0) f(x) + 1 - P_{\text{ava}}(0)}, \tag{2}$$

$$P_{\text{ava}}(0) = 1 - \exp\left[-\int_0^L \alpha_h P_{\text{ava}}(x') dx'\right], \tag{3}$$

$$f(x) = \exp\left(-\int_0^x \alpha_e - \alpha_h dx'\right). \tag{4}$$

Here, L is the length of the depletion layer. $P_{\text{ava}}(x)$ and $P_{\text{ava}}(0)$ mean the avalanche probabilities of a single carrier at x and 0. By adjusting the bias voltage, we change the density of carrier at $x = 0$. That is how the voltage influences the avalanche current.

Collisional ionization theory gives us a systematic approach to modeling the avalanche in Si diodes. Now, we consider the random source in our experiment—the dark count effect in Si diodes. Thermal noise and tunneling-generation noise are the two main factors that contribute to dark counts.

Thermal noise is determined by the intrinsic properties of materials and working temperature of diodes:

$$G_{Ti} = \frac{n_i}{\tau_i}. \tag{5}$$

The i's are labels of different layers: n_i represents the number density of the intrinsic carrier in layer i and τ_i is the lifetime of free carriers in layer i. n_i and τ_i are both functions of temperature. Thermal noise signals are suppressed in our experiment by a cooling system.

Tunneling-generation noise signals are more complicated than thermal noise. They include band-band-tunneling (BBT) signals [29] and trap-assisted-tunneling (TAT) signals [30]. Namely, free carriers could tunnel directly through the band gap or take a two-step tunneling by first jumping to a trapped state. Quantitatively, the number of free carriers generated by these two processes in diodes can be considered in the following way [31]:

$$N_i = \frac{J_{\text{BBT}} + J_{\text{TAT}}}{q}, \tag{6}$$

$$J_{\text{BBT}} = \sqrt{\frac{2m_r}{E_g}} \frac{q^2 F^2}{4\pi^3 \hbar^2} \exp\left(-\frac{\pi\sqrt{m_r E_g^3}}{2\sqrt{2} q\hbar F}\right), \tag{7}$$

$$J_{\text{TAT}} = \frac{\sqrt{\frac{2m_r}{E_g}} \frac{q^2F^2}{4\pi^3\hbar^2} N_{\text{trap}} \exp\left(-\frac{\pi\sqrt{m_{lh}E_{B1}^3}+\pi\sqrt{m_c E_{B2}^3}}{2\sqrt{2}q\hbar F}\right)}{N_\upsilon \exp\left(-\frac{\pi\sqrt{m_{lh}E_{B1}^3}}{2\sqrt{2}q\hbar F}\right) + N_c \exp\left(-\frac{\pi\sqrt{m_c E_{B2}^3}}{2\sqrt{2}q\hbar F}\right)}. \tag{8}$$

Here, BBT represents band-band tunneling; TAT represents trap-assisted tunneling; q is carrier charge; m_r is

effective mass; E_g is band gap; and F is electric field, which is a function of position. N_{trap}, N_υ, and N_c are the density of traps and density of light-hole states in the valance and conduction band, respectively. m_{lh} and m_c are the effective mass of the light hole and electron in the conduction band. E_{B1} and E_{B2} are the energy gaps between the valance band and trap and between the trap and conduction band, respectively. This part of the dark count signals is utilized as a random source in our experiment.

The dark count rate (DCR) of a diode is the integration of G_{Ti} and N_i through the whole absorption layer and depletion layer:

$$\text{DCR} = \int (G_{Tdep} + N_{dep}) P_p(x')\, dV_{\text{dep}} + P_p(0) \int (G_{Tab} + N_{ab})\, dV_{ab}. \tag{9}$$

DCR give us the rate of dark signals in the Si diode. Furthermore, considering the practical situation, we assume G_{Ti} to be far smaller than N_i and α_e and α_h to be constants under a fixed bias voltage.

Based on these discussions, we prepare our system in the following way to make sure that tunneling effects dominate the whole process:

(1) The optical input port is turned off to prevent photon absorption.

(2) Also, we add a cooling system to make sure the temperature of the working environment is low enough that the thermal excitation of the free carriers can be neglected. In our experiment, the system is cooled to -20°C.

(3) Furthermore, we apply an active-quenching system to decrease the afterpulse effect. We set the hold-off time after to 17 ns.

Yet, we still cannot eradicate noise signals without a preselection process.

In our experiment, we focus more on the property of random numbers generated by current devices rather than the optimization of the experiment system. Thus, for simplicity, we choose a Si SPAD from a Si single-photon detector called SPD4, produced by ROI Optoelectronics Technology Co., Ltd. We do not change the inner structure of it in the following experiments. Moreover, this commercialized product gives us a stable Peltier cooling system. It maintains our system at -20°C with a possible deviation of less than 0.2°C in a 4-h consecutive operation.

III. SCHEME

Previous studies on collisional ionization demonstrated the motion of free carriers in Si diodes. Tunneling signals are a major part of dark counts in the low-temperature case. When we apply a proper bias voltage to Si SPAD, electrons inside tunnel through the valence band and conduction band with a certain probability and trigger avalanche signals. One direct way to extract randomness is to apply a periodic signal to diodes and then write down "1" when a tunneling signal is detected in one period or "0" when there is no tunneling event in the time period of the trigger signal. But this method requires a high frequency and stable power source. Therefore, in order to simplify the experiment as well as cut down cost, we apply a dc power to the diode, which is enough to trigger tunneling signals and then take the time intervals of the adjacent tunneling signals as a random source, as shown in Fig. 1.

The basic thought of our experiment is that, under a certain bias voltage, the dark count rate of tunneling signals in Si diodes is almost a constant. These tunneling signals are amplified by the avalanche effect inside the Si diode. The exact time of when the tunneling events take place is completely stochastic. Random numbers are generated from these time intervals. The detailed scheme can be expressed as follows:

Step 1 We first cool the Si SPAD down to -20°C and apply a dc power with a constant voltage on it. This process is to reduce the thermal dark counts and make sure that the tunneling dark count rate is stable during the random-number generation process.

Step 2 Connect the signal collection module and a standard clock to detect the tunneling signals. In one period of the clock, tunneling probability is a constant as the voltage is fixed. We write down the number of periods between two adjacent tunneling signals. Then, we write down how many tunneling events are recorded in 1 s. These numbers are vital to our analysis later.

Step 3 Connect the data collection module to a PC and run the data collection software to record the number of tunnelings in a second. Data are recorded into a text file.

Step 4 After the generation speed is stabilized, start data collection.

Step 5 Turn off the whole system, preselect the data and conduct a randomness extraction.

Step 6 Check the randomness of the final data with standard tests.

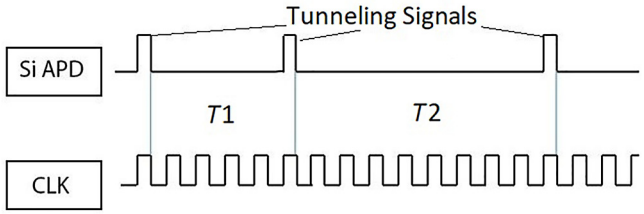

FIG. 1. Simplified diagram of how we record the time intervals. APD, avalanche photodiode; CLK, clock.

As we measure the time intervals of adjacent tunneling signals by a standard clock, whose frequency is 500 MHz, we assume that in one period of this clock (2 ns), the probability of electron tunneling is p and very small.

The voltage and temperature are fixed, so we assume that p here is a constant. We denote it as p_0.

Assuming that tunneling signals are independent of one another, we can directly come to the probability that the time interval $t = n \times T$ between adjacent signals under a fixed voltage U_0:

$$\begin{aligned} P(n) &= (1-p_0)^n p_0 \\ &= p_0(1-p_0)^{n/p_0 p_0} \\ &= p_0 e^{-np_0}. \end{aligned} \tag{10}$$

Here, T refers to the period of a standard clock, so the probability obeys an exponential distribution, as shown in Fig. 2. However, we have to consider the afterpulse effect in the Si diode [32]. Every single tunneling signal leaves electron-vacancy pairs in the diode and thus enhances the tunneling probability in the next few clock periods. For a sufficiently long time after the former tunneling signal or another tunneling event happens, these electron-vacancy pairs disappear so the tunneling probability should go back to p_0. With this assumption, we can rewrite the probability as

$$P(n) = [p_0 + p_a(t_n)] \prod_{i=1}^{i=n-1} [1 - p_0 - p_a(t_i)]. \tag{11}$$

$p_a(t_i)$ refers to the additional probability brought by the afterpulse effect of a tunneling signal after i clock periods. In the next section, we conduct our data preselection based on the equation above.

Before we make a detailed analysis of $p_a(t_i)$, we first collect the 60-MB output of the Si SPAD under 98 V and draw a dark count time diagram with theoretical results versus experimental results as shown below:

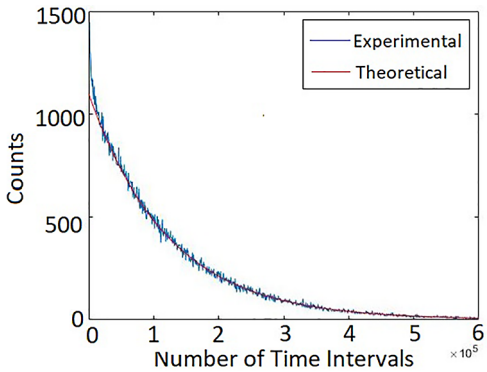

FIG. 2. Theoretical prediction vs experimental results under 98-V, 60-MB data.

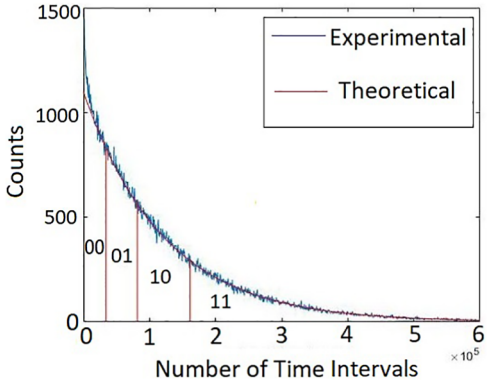

FIG. 3. Simple concept of preliminary encoding on output data.

Except for the difference at $n = 5 \times 10^4$, the experimental result and theoretical prediction are highly fitted. Thus, we can transform real signals into a uniform-distributed data. We divide the area covered by the theoretical line into 1024 equal fractions. Then, we encode all the experimental data in each of them with the binary form of the faction's serial number. The basic principle is as shown in Fig. 3. We apply this process to 1.5 GB data collected under 100 V and get Fig. 4.

As we can see from Fig. 4, counts of random numbers near 0 are much more than theoretical prediction. Thus, we have to take $p_a(t_i)$ into consideration to eliminate this difference. This process is illustrated in the next section.

IV. PRESELECTION AND POSTPROCESSING

In this section, we introduce the preselection and postprocessing processes. As stated previously, we have four parts in our Si quantum random-number generator: the Si SPAD part, power and cooling system, active-quenching system, and data-processing system.

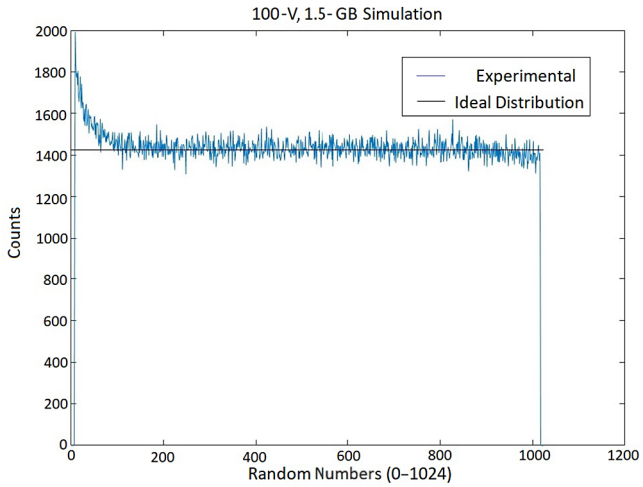

FIG. 4. Results of data after encoding.

We exploit a semiconductor cooling system to make sure the working temperature of Si SPAD is −20°C. This system is stable after 4 h of consecutive running.

Aside from the thermal noise, another adversary of our system is the afterpulsing effects. As we illustrated before, each tunneling signal leaves a few remaining free carriers in the diode, captured by the inner defects of diodes. These captured carriers are released later. This effect increases the probability of detecting tunneling signals. Our job now is to reduce the influence of this effect.

The magnitude of the afterpulse is associated with several factors. The probability of afterpulse can be expressed as

$$P_a \propto CT \exp\left(-\frac{t}{\tau}\right). \tag{12}$$

C is the sum of the Si SPAD's effective capacitance and the parasitic capacitance of the whole circuit, which is a constant. T refers to the duration of each avalanche. t is the "hold-off time," here 17 ns. τ is the lifetime of the free carriers, which is also a constant. Once our system is set, P_a is an exponential to t. We can further simplify it to

$$P_a(t) = A \exp(-Bt). \tag{13}$$

A is a constant coefficient and B is $1/\tau$. In order to reduce the afterpulse effect, researchers have designed different systems. Basically, there are three main methods: passive quenching, active quenching, and gate-pulse quenching [33].

Each of these methods has advantages and drawbacks. In our system, we choose the active-quenching circuit. Namely, after each avalanche event, the voltage on SPAD is lowered to 0 V for 17 ns to decrease the afterpulse effect. This circuit protects the Si SPAD from the consecutive Geiger mode and has a fast response time. However, the on-off of the switch would cause sharp noises and it would take a more complex circuit to reduce such noises, which is not elaborated here.

Now, we take the logarithm of Eq. (11) and get the following equation (in this paper, we use "ln" to represent the natural logarithm):

$$\ln[P_r(n)] = \ln[p_0 + P_a(t)] + \sum_{i=1}^{n-1} \ln[1 - p_0 - P_a(t_i)]. \tag{14}$$

Here, n means the number of periods counted by a standard clock. The periods are also quite short, 2 ns. We can use integration to substitute the summation in Eq. (14):

$$\ln[P_r(t)] = \ln[p_0 + P_a(t)] + \alpha \int_0^t \ln[1 - p_0 - P_a(t')]\, dt'. \tag{15}$$

Furthermore, we have $P_a(t) \ll p_0 \ll 1$:

$$\ln[P_r(t)] = \ln[p_0 + P_a(t)] - \alpha \int_0^t [p_0 + P_a(t')] dt' \tag{16}$$

$$P_r(t) = [p_0 + P_a(t)] \exp\left\{-\alpha \int_0^t \left[p_0 + P_a(t')\right]\, dt'\right\}. \tag{17}$$

Here, α is a constant that represents the time density of probability. In order to test our assumption that the afterpulse effect adds an additional exponential probability to later tunneling signals, we calculate the quotient of $P_r(t)$ on $P(t)$:

$$\frac{P_r(t)}{P(t)} = \frac{p_0 + P_a(t)}{p_0} \exp\left[-\alpha \int_0^t P_a(t')\, dt'\right]. \tag{18}$$

Then we calculate the logarithm of this quotient:

$$\ln\left(\frac{P_r(t)}{P(t)}\right) = \ln\left(\frac{p_0 + P_a(t)}{p_0}\right) - \alpha \int_0^t P_a(t')\, dt'. \tag{19}$$

Considering the fact that $P_a(t) \ll p_0$, we bring the concrete expression into Eq. (19):

$$\begin{aligned} \ln\left(\frac{P_r(t)}{P(t)}\right) &= \frac{A\alpha \exp(-Bt)}{p_0} + \frac{A\alpha}{B} \exp(-Bt) \\ &= \frac{A\alpha(p_0 + B)}{p_0 B} \exp(-Bt). \end{aligned} \tag{20}$$

So under the assumption that $P_a(t) \ll p_0 \ll 1$, this logarithm of the quotient should be an exponential of t. Then, we calculate this value for output data under 94, 98, 100, and 102 V, respectively. Here, for simplicity, we only show data under 102 V, as shown in Fig. 5.

We now find out that the experimental results fit the theoretical prediction under all these bias voltages. Furthermore, according to Eq. (20), here $\tau = 5.3$ μs. The average time interval between two adjacent tunneling signals is 149 μs. Thus, we can say that the influence brought by the afterpulse effect is short ranged and we can correct the deviation based on our assumption. Taking this noise model into consideration, we cut the number of counts at different n values by a specific proportion. We call this preselection before postprocessing. We, again, preselect the 1.5-GB data under 100 V and then take a part of it to show the distribution. The result is shown in Fig. 6.

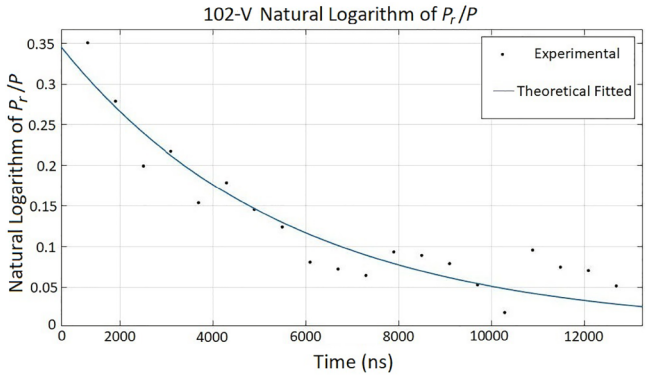

FIG. 5. Logarithm of quotient under 102 V.

So far, we have modified the output by the assumption that noise generated by the afterpulse effect is exponentially decaying.

One of the most important parameters in postprocessing, min-entropy [34], is also optimized after preselection. Randomness extractors based on min-entropy have been widely used in previous work on QRNGs [35–37]. In our scheme, we use the Toeplitz-hashing extractor for randomness extraction [38]. For a sequence consisting of random variable X, the definition of min-entropy is

$$H_m = -\log_2(\max X). \tag{21}$$

Min-entropy partially represents the randomness of the sequence. The more chaotic the sequence is, the higher the min-entropy. Extractors are supposed to distill the randomness in X and omit the rest. They often work in the following pattern:

$$\{0,1\}^m \bigotimes \text{Seed} \xrightarrow{\text{extract}} \{0,1\}^n. \tag{22}$$

The Toeplitz-hashing extractor and Trevisan extractor are both widely used extractors. Here, we choose the former. In the Toeplitz-hashing extractor, we have to build a Toeplitz matrix based on the min-entropy and then calculate the product of the output data and this matrix to get the final data.

In our experiment, we calculate the min-entropy for each 10 bits of our output data. After preselection, the min-entropy is 9.79. This result means that we can preserve almost 98% of data after the postprocessing, which indicates that our system is quite efficient. We test the final data with Diehard and NIST in the next section.

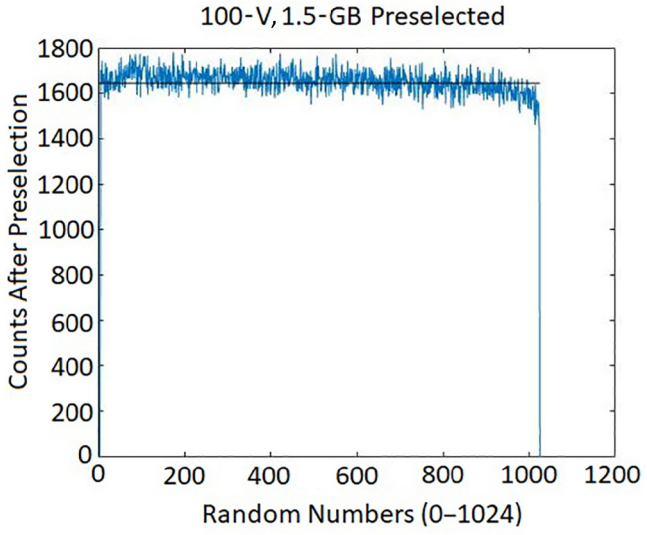

FIG. 6. Distribution of preselected random number.

V. RANDOMNESS TEST

After postprocessing, we finally get 1.5-GB data with a 100-V bias voltage. The next step is the randomness test. There are several kinds of randomness tests. For instance, the NIST Test suite [21], Diehard test [22], and TestU01 [23]. Each of them contains subsections of detailed statistical tests, verifying the randomness of input data from many different aspects.

Here, we use the NIST test and Diehard test to check the randomness of final data.

The criterion for passing NIST is that the p-value of each test lies between 0.01 and 0.99. Our data meet this criterion with a passing rate of 98.2%. The criterion of passing the Diehard test is the p-value lies between 0.000001 and 0.999999. Concrete results are listed in the Appendix.

As shown, our final data pass these two tests. Another advantage is that all the information needed in our data encoding, preselection, and postprocessing is actually only the average number of tunneling pulses in 1 s, which is controlled by the bias voltage. This parameter shows up when we are trying to divide the area into 1024 parts.

Thus, we implement a Si QRNG based on tunneling effects in Si diodes under a dc input. Its data output speed is 6.98 MB/s. This system, with a quite simple and cheap setup, can work stably for more than 4 h.

VI. CONCLUSION

This paper introduces our work on the practical design of a Si QRNG based on tunneling effects. Our system is quite simple and cheap in comparison with many other QRNG systems with a photon source. The performance of our system can also be enhanced by improving the hardware of each module, which will be studied in our future work.

There are some other aspects that future work might focus on. Although the output speed of our system could

be enhanced, it cannot exceed most continuous QRNG systems. This outcome means that our system could not serve in many high-speed circumstances. It is not the problem of the tunneling effects but the Si diode system. Another system should be studied to achieve a higher generation speed.

Another problem is that the deviation caused by afterpulse noise is illustrated but not efficiently solved. The reverse procedure of cutting this deviation down is not that easy. We have not found a method to treat these data as an ensemble. This problem leads to the difficulty of integrating this preselection part into a FPGA. For now, this preselection part is implemented using MATLAB on PC.

In conclusion, our work proves the possibility of setting up a cheap, practical, photon-free, and stable QRNG. However, there are still many aspects to be optimized. We hope our work can give some useful ideas in designing practical QRNGs.

ACKNOWLEDGMENTS

We thank Pan Dong, Gao Xingyu, and Liu Yipu for their useful discussions on the structure of Si SPAD. We also thank Liu Xinyu, Jiang Nan, and Ai Fei for their assistance in designing and using the software. Support from Beijing Advanced Innovation Center for Future Chip (ICFC) is gratefully acknowledged. We also appreciate financial support from the National Nature Science Foundation of China (Grant No. 04130211). Finally, we appreciate the suggestions from Professor David Aspnes.

TABLE I. Results of the NIST test for the 1.5-GB final data. The minimum pass rate for each statistical test with the exception of the random excursion (variant) test is approximately 1499 for a sample size of 5160 binary sequences. The minimum pass rate for the random excursion (variant) test is approximately 517 for a sample size of 524 binary sequences. With a confidence parameter $\alpha = 0.01$, our data pass the NIST test.

Statistical test	p-value	Proportion	Assessment
Frequency	0.761328	0.990119	Success
Block frequency	0.874053	0.990514	Success
Cumulative sums	0.897521	0.989526	Success
Runs	0.979283	0.994664	Success
Longest run	0.345221	0.990316	Success
Rank	0.102931	0.992688	Success
FFT	0.764302	0.989723	Success
Nonoverlapping template	0.223218	0.987945	Success
Overlapping template	0.298453	0.990316	Success
Universal	0.496539	0.987945	Success
Approximate entropy	0.978821	0.992292	Success
Random excursions	0.622942	0.993312	Success
Random excursions variant	0.432522	0.986641	Success
Serial	0.710092	0.989723	Success
Linear complexity	0.447382	0.992095	Success

TABLE II. Results of the Diehard test for the 1.5-GB final data. All of these indexes lie within (0.000001, 0.999999), so our data pass the Diehard test.

Statistical test	p-value	Assessment
Birthday test	0.498372	Success
Overlapping permutation	0.387202	Success
Ranks of 31 × 31 matrices	0.887632	Success
Ranks of 32 × 32 matrices	0.342685	Success
Ranks of 6 × 8 matrices	0.912761	Success
Bitstream test	0.102923	Success
OPSO	0.6521	Success
OQSO	0.8467	Success
DNA	0.7129	Success
Count 1 s in the stream of bytes	0.273583	Success
Count 1 s in the special bytes	0.844794	Success
Parking lot test	0.247862	Success
Minimum distance test	0.9221	Success
3D spheres test	0.776623	Success
Squeeze test	0.386842	Success
Overlapping sums test	0.299432	Success
Runs	0.756887	Success
Craps	0.194702	Success

APPENDIX: DETAILED RESULT OF RANDOMNESS TESTS

A detailed data analysis using the NIST test is obtained by the official program STS version 2.1.2, as shown in Table I. A detailed data analysis using the Dihard test is shown as the following Table II.

[1] C. Robert, Machine learning, a probabilistic perspective, Chance **27**, 62 (2014).
[2] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. A. Smolin, Experimental quantum cryptography, Theory Appl. Cryptographic Tech. **5**, 3 (1991).
[3] J. P. Paz, Randomness in quantum computation, Science **302**, 2076 (2003).
[4] F. G. Deng and G. L. Long, Secure direct communication with a quantum one-time pad, Phys. Rev. A **69**, 052319 (2004).
[5] S. K. Park and K. W. Miller, Random number generators: Good ones are hard to find, Commun. ACM **31**, 1192 (1988).
[6] J. Soto and L. E. Bassham, *Randomness Testing of the Advanced Encryption Standard Finalist Candidates* (Booz-Allen and Hamilton Inc., Mclean VA, 2000).
[7] T. Erber and S. Putterman, Randomness in quantum mechanics—nature's ultimate cryptogram? Nature **318**, 41 (1985).

[8] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, A fast and compact quantum random number generator, Rev . Sci. Instrum. **71**, 1675 (2000).

[9] M. A. Wayne, E. Jeffrey, G. M. Akselrod, and P. G. Kwiat, Photon arrival time quantum random number generation, J. Mod. Opt. **56**, 516 (2009).

[10] H. Q. Ma, Y. J. Xie, and L. A. Wu, Random number generation based on the time of arrival of single photons, Appl. Opt. **44**, 7760 (2005).

[11] F. H. Xu, B. Qi, X. F. Ma, H. Xu, H. X. Zheng, and H. K. Lo, Ultrafast quantum random number generation based on quantum phase fluctuations, Opt. Express **20**, 12366 (2012).

[12] B. Qi, Y. M. Chi, H. K. Lo, and L. Qian, High-speed quantum random number generation by measuring phase noise of a single-mode laser, Opt. Lett. **35**, 312 (2010).

[13] Y. Q. Nie, L. L. Huang, Y. Liu, F. Payne, J. Zhang, and J. W. Pan, The generation of 68 gbps quantum random number by measuring laser phase fluctuations, Rev. Sci. Instrum. **86**, 063105 (2015).

[14] T. Lunghi, J. B. Brask, C. C. Lim, Q. Lavigne, J. Bowles, A. Martin, H. Zbinden, and N. Brunner, Self-testing Quantum Random Number Generator, Phys. Rev. Lett. **114**, 150501 (2015).

[15] Z. Cao, H. Y. Zhou, X. Yuan, and X. F. Ma, Source-independent Quantum Random Number Generation, Phys. Rev. X **6**, 011020 (2016).

[16] Y. Bomze, Y. Hey, H. T. Grahn, and S. W. Teitsworth, Noise-induced Current Switching in Semiconductor Superlattices: Observation of Nonexponential Kinetics in a High-dimensional System, Phys. Rev. Lett. **109**, 026801 (2012).

[17] W. Li, I. Reidler, Y. Aviad, Y. Y. Huang, H. L. Song, Y. H. Zhang, M. Rosenbluh, and I. Kanter, Fast Physical Random-number Generation Based on Room-temperature Chaotic Oscillations in Weakly Coupled Superlattices, Phys. Rev. Lett. **111**, 044102 (2013).

[18] H. Zhou, J. Li, D. Pan, W. Zhang, and G. Long, Quantum random number generator based on quantum tunneling effect. ArXiv e-prints, 17110.1752Z, (2017).

[19] L. E. Tarof, D. G. Knight, K. E. Fox, C. J. Miner, N. Puetz, and H. B. Kim, Planar inp/ingaas avalanche photodetectors with partial charge sheet in device periphery, Appl. Phys. Lett. **57**, 670 (1990).

[20] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, Evolution and prospects for single-photon avalanche diodes and quenching circuits, J. Mod. Opt. **51**, 1267 (2004).

[21] A. Rukhin, J. Soto, J. Nechvatal, S. Miles, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, and A. Heckert, A statistical test suite for random and pseudorandom number generators for cryptographic applications, Appl. Phys. Lett. **22**, 1645 (2010).

[22] G. Marsaglia, *The Marsaglia Random Number CDROM Including the Diehard Battery of Tests of Randomness* (Florida State University, Florida, 1995).

[23] P. L'Ecuyer and R. Simard, Testu01: A c library for empirical testing of random number generators, ACM Trans. Math. Software **33**, 22 (2007).

[24] B. Julian and R. Anthony, US Patent No. 2,849,664 (1958).

[25] W. G. Spitzer and H. Y. Fan, Determination of optical constants and carrier effective mass of semiconductors, Phys. Rev. **106**, 882 (1957).

[26] R. N. Noyce, US Patent No. 2,981,877 (1961).

[27] R. J. Mcintyre, Multiplication noise in uniform avalanche diodes, IEEE Trans. Electron Devices **13**, 164 (1966).

[28] R. J. Mcintyre, The distribution of gains in uniformly multiplying avalanche photodiodes: Theory, IEEE Trans. Electron Devices **19**, 703 (1972).

[29] A. Schenk, Rigorous theory and simplified model of the band-to-band tunneling in silicon, Solid-State Electron. **36**, 19 (1993).

[30] M. O. Andersson, Z. Xiao, S. Norrman, and O. Engström, Model based on trap-assisted tunneling for two-level current fluctuations in submicrometer metal–silicon-dioxide–silicon diodes, Phys. Rev. B. **41**, 9836 (1990).

[31] J. P. Donnelly, E. K. Duerr, K. A. Mcintosh, E. A. Dauler, D. C. Oakley, S. H. Groves, C. J. Vineis, L. J. Mahoney, K. M. Molvar, and P. I. Hopman, Design considerations for 1.06-μm ingaasp–inp geiger-mode avalanche photodiodes, IEEE J. Quantum Electron. **42**, 797 (2006).

[32] A. D. Mora, D. Contini, A. Pifferi, R. Cubeddu, A. Tosi, and F. Zappa, Afterpulse-like noise limits dynamic range in time-gated applications of thin-junction silicon single-photon avalanche diode, Appl. Phys. Lett. **100**, 241111 (2012).

[33] A. Lacaita, C. Samori, F. Zappa, M. Ghioni, and S. Cova, Avalanche photodiodes and quenching circuits for single-photon detection, Appl. Opt. **35**, 1956 (1996).

[34] R. A. Wiggins, Minimum entropy deconvolution, Geophys. Prospect. Petrole **16**, 21 (1980).

[35] O. Chevassut, P. A. Fouque, P. Gaudry, and D. Pointcheval, Key derivation and randomness extraction, IACR Cryptology ePrint Archive **2005**, 61 (2005).

[36] E. Kiltz, K. Pietrzak, M. Stam, and M. Yung, in *International Conference on Advances in Cryptology: The Theory and Applications of Cryptographic Techniques* (2009), p. 590.

[37] Y. Dodis, A. Elbaz, R. Oliveira, and R. Ran, in *International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, Approx 2004, and International Workshop on Randomization and Computation, Random 2004, Cambridge, MA, USA, August 22–24, 2004, Proceedings* (2004), p. 334.

[38] X. F. Ma, F. H. Xu, H. Xu, X. Q. Tan, B. Qi, and H. K. Lo, Postprocessing for quantum random number generators: Entropy evaluation and randomness extraction, Phys. Rev. A. **87**, 062327 (2013).