Finite-size analysis of continuous variable source-independent quantum random number generation

Junyu Zhang¹ · Yichen Zhang¹ · Ziyong Zheng¹ · Ziyang Chen² · Bingjie Xu³ · Song Yu¹

Received: 19 February 2020 / Accepted: 10 November 2020 / Published online: 7 January 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract
We study the impact of finite-size effect on continuous variable source-independent quantum random number generation. The central limit theorem and maximum likelihood estimation theorem are used to derive the formula which could output the statistical fluctuations and determine upper bound of parameters of practical quantum random number generation. With these results, we can see the check data length and confidence probability has intense relevance to the final randomness, which can be adjusted according to the demand in implementation. Besides, other key parameters, such as sampling range size and sampling resolution, have also been considered in detail. It is found that the distribution of quantified output related with sampling range size has significant effects on the loss of final randomness due to finite-size effect. The overall results indicate that the finite-size effect should be taken into consideration for implementing the continuous variable source-independent quantum random number generation in practical.

Keywords Finite-size effect · Quantum random number generation · Statistical fluctuations

Yichen Zhang
zhangyc@bupt.edu.cn

¹ State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
² State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronics, and Center for Quantum Information Technology, Peking University, Beijing 100871, China
³ Science and Technology on Security Communication Laboratory, Institute of Southwestern Communication, Chengdu 610041, China
1 Introduction

Random numbers are traditionally generated using deterministic processes. However, these predictable random numbers will lead to security loopholes or errors in some fields, such as cryptography, scientific simulation and lottery. As an ideal alternative, quantum random number generators (QRNGs) have attracted much attention in the past few years [1,2] because they exploit the fundamental indeterminacy of quantum mechanics and can generate unpredictable random numbers. Based on the randomness reliability levels, three types can be categorized in QRNGs: practical QRNGs, device-independent QRNGs and semi-device-independent QRNGs. Practical QRNGs [3–15] completely trust their well-characterized devices and generate random numbers with high speed. Device-independent (DI) QRNGs [16,17] have the most paranoid assumption that the devices are all untrusted and have the highest security levels, which results to low generation speed that can hardly meet actual demands. As a trade-off, semi-device-independent QRNGs generate high-speed random numbers by just trusting parts of the devices. One mainly considers QRNGs with untrusted measurement devices [18,19] or untrusted sources in this field. Therein, source-independent (SI) QRNGs get people’s favor due to the difficulty of pure source state preparation [20–25].

Imparity with low-speed discrete variable SI-QRNGs [26], continuous variable (CV) SI-QRNGs achieve higher speed. CV-SI-QRNGs eliminate the effects of malicious eavesdropper through randomly switching the measurement bases generally with phase modulator. The application of phase-randomized local oscillator with gain-switched laser makes it easier to implement [21]. The heterodyne measurement remarkably increases the generating speed to 17 Gbps and operates without an initial source of randomness [24]. Several practical imperfections have been considered in [23], and the generating speed achieves 15.07 Gbps high generation speed. In most CV-SI-QRNG protocols, the mutual information between Alice and Eve is the most crucial parameter, which should be estimated with generated raw sequences in practice. To guarantee the security, CV-SI-QRNGs always discard untrusted parts of information by post-processing algorithms, such as Toeplitz hashing matrix, after parameter estimation.

The estimation of precise parameters requires infinite size of check data. However, the size of check data in practice is finite, because any practical system can only run in finite time. This will lead to the statistical fluctuations of estimated parameters, i.e., we always overestimate or underestimate the parameters compared with their ideal values. This may result in miscalculating the available randomness and lead to security loopholes. Thus, the finite-size effect is necessary to be considered in practical systems. Some important previous works have been made in this field. Reference [22] uses smooth min-entropy to estimate the loss of entropic uncertainty and calculate the lower bound of it. Reference [20] uses the composable security definition to consider finite size effect to bound the phase error rate and then estimate the upper bound of the number of bits sacrificed in the phase error correction. It is different from these two references that we propose a new method, which consider central limit theorem and the impact of some key parameters in detail.
We study finite-size effect on the CV-SI-QRNG protocol [23] by calculating the statistical fluctuations of estimated parameters. The central limit theorem and maximum likelihood estimation theorem are utilized to find the distribution of estimated parameters. The statistical fluctuation of final randomness can be computed out with a given confidence probability. Then, we give numerical simulations on extractable randomness with different variable parameters including check data length \( m \), confidence probability \( \epsilon \), sampling range size \( N \) and sampling resolution \( n \) for studying the impacts of finite-size effect under various scenarios. Finally, we analyze the relationship between sampling range size and finite-size effect in detail.

In this paper, we introduce our work as follows: In Sect. 2 we review the CV-SI-QRNG protocol proposed in [23]. In Sect. 3 we briefly describe the security analysis of our protocol, then consider the effect of finite size on the security of the protocol in detail and derive the formulas for the analysis of finite size. In Sect. 4 we show and discuss the results of simulation with several variable parameters which could impact the finite-size effect. At last, we give conclusions in Sect. 5.

2 Continuous variable source-independent QRNG protocol

The more detailed introduction about the CV-SI-QRNG protocol is in [23]; we just make a brief introduction here. Figure 1 illustrates the schematic of our CV-SI-QRNG protocol. There is an untrusted and uncharacterized source prepared by Eve with quantum state \( \rho_A \), which is allowed in infinite dimensions. Eve keeps a quantum system \( \bar{E} \) related to system \( A \) and then sends \( \rho_A \) to Alice for homodyne measurement. To extract security randomness, Alice switches the measurement bases randomly between \( X \) and \( P \) quadratures for measuring \( \rho_A \) with an initial random seed.

Obviously, the switch should be in an independently and identically distribution (i. i. d.) way to extract i. i. d. random numbers. Now we define \( n_{\text{tot}} \) as the number of total measurement result, \( n_c \) as the number of check samples out of total measurements and \( t \) as the length of random seed. It should be noted that we make an assumption that the measurement can be trusted and calibrated well; meanwhile, a quantum correlated Eve can introduce all the excess noise.

After Alice measures the \( X \) and \( P \) quadratures and gets check sequence, the covariance matrix (CM) of the state \( \rho_A \) can be estimated with this formula:

\[
\gamma_A = \begin{bmatrix} V_x & c \\ c & V_p \end{bmatrix},
\]

where \( V_x \) and \( V_p \) are the variance of \( X \) and \( P \) quadratures, and \( c \) is the covariance between \( X \) and \( P \) quadratures for \( \rho_A \).

At last, the final randomness that Alice could extract from total \( n_{\text{tot}} \) measurement is asymptotically calculated by \( (n_{\text{tot}} - n_c) (H (a_i) - S (a_i : E)) - t \), where \( H (a_i) \) is the discrete Shannon entropy of Alice’s measurement results and \( S (a_i : E) \) is the quantum mutual information between the quantum state \( \rho_{\bar{E}} \) kept by Eve and Alice’s measurement results. Then putting the remaining measurement results into randomness extractor [27,28]. There are several mature and verifiable randomness
Fig. 1 Schematic of CV-SI-QRNG protocol. Eve completely controls the untrusted source and keeps quantum state $\rho_E$ related with quantum state $\rho_A$ which is sent to Alice. Then, Alice does balanced homodyne measurement on $\rho_A$ with bases which switched randomly between $X$ and $P$ quadratures. The results of measurement $a_i$ must be eliminated the insecurity parts by post-processing to output available security random numbers extractors that can eliminate the untrusted part of randomness and remain secure true randomness based on the parameters set by Alice.

In the whole protocol we introduced above, no assumptions have been made on the source of QRNG. That means, the random numbers extracted are secure enough regardless of the purity and dimensions of Alice’s input state. The measurement devices have our fully trust since it is inaccessible for Eve.

3 Security analysis and finite-size effect

In this section, we will make a brief analysis about the security of our CV-SI-QRNG protocol firstly (the more detailed analysis is in [23]) and the formulas that are used in finite-size analysis and numerical simulations are derived.

3.1 Security analysis

Our protocol is very similar with continuous variable quantum key distribution [29–31], the key rate of our CV-SI-QRNG is calculated with the Devetak–Winter formula [32], in the case of i. i. d, which is used in continuous variable quantum key distribution [33,34]. The definition is given by:

$$K = \beta (I(a : b) - S(a : E)),$$

where $\beta$ is the reconciliation efficiency, $I(a : b)$ is the classical mutual information between Alice and Bob and $S(a : E)$ is the Holevo bound between the quantum state $\rho_E$ kept by Eve and the result $a$ measured by Alice. Different from continuous variable quantum key distribution scenario, in QRNG, Alice and Bob are at the same station. Thus, we do not perform information reconciliation. Therefore, the mutual information $I(a : b)$ is described by Shannon entropy $H(a)$ which is just defined by Alice’s actually obtained data and the reconciliation efficiency $\beta$ equals to 1.

In Eq. 2, it should be noted that Alice’s measurement result $a$ is a continuous variable, that means the homodyne detection is perfect with infinite range and resolution. But, in practical, the measurement of Alice must be imperfect with finite range and
finite resolution, i.e., it should be coarse gained and discrete [35]. It mainly depends on the precision of analog-to-digital convertor, which digitizes the unreadable continuous data \( a \) into \( n \) bit \( a_i \). We define the sampling range of the analog-to-digital convertor is \([-N + \Delta/2, N - \Delta/2]\), where \( \Delta = N/2^{n-1} \). The sampling range has been divided evenly into \( 2^n \) bins and \( \Delta \) is the interval between every adjacent \( a_i \).

Then, the available secure randomness that uncorrelated with state \( \rho_E \) kept by Eve can be calculated out in our condition by formula:

\[
R_{\text{dis}}(a_i | E) = H(a_i) - S(\rho_E | a_i : E),
\]

(3)

where \( H(a_i) \) is the discrete Shannon entropy of \( a_i \), and \( S(\rho_E | a_i : E) \) is the Holevo bound between the quantum state \( \rho_E \) kept by Eve and the discrete result \( a_i \) measured by Alice. \( H(a_i) \) can be calculated easily with Alice’s measurement result \( a_i \). So, we focus on \( S(\rho_E | a_i : E) \) to find its upper bound due to the difficulty of calculating it out directly.

First, there is an inequality proved in [23] that \( S(\rho_E | a_i : E) \leq S(a : E) \). It means the digitization operation will not increase the mutual information between the quantum state \( \rho_E \) kept by Eve and the discrete result \( a_i \) measured by Alice. Then, we consider the worst case that the state \( \rho_E \) kept by Eve is a purification of \( \rho_A \). This leads to \( S(\rho_E | a) = 0 \) and \( S(a : E) = S(\rho_E) = S(\rho_A) \). The von Neumann entropy of Gaussian state \( \rho_A^G \) is the upper bound in that of arbitrary state \( \rho \) if they have the same CM [36]. So, there is \( S(\rho_A) \leq S(\rho_A^G) \).

Now, the lower bound of final extractable randomness [37] goes to

\[
R_{\text{dis}}(a_i | E) \geq H(a_i) - S(\rho_A^G). \tag{4}
\]

Focusing on \( S(\rho_A^G) \), its von Neumann entropy is

\[
S(\rho_A^G) = \frac{\lambda + 1}{2} \log_2 \frac{\lambda + 1}{2} - \frac{\lambda - 1}{2} \log_2 \frac{\lambda - 1}{2}, \tag{5}
\]

where \( \lambda \) is the symplectic eigenvalue of matrix in Eq. 1 and \( \lambda = \sqrt{\det(\gamma_A)} = \sqrt{V_x V_p - c^2} \). But we cannot obtain the exact value of \( \lambda \) due to the loss information inside the digitized interval. Therefore, we have to estimate its upper bound to obtain secure randomness. We set \( c = 0 \) and defined \( \lambda < \tilde{\lambda} = \sqrt{\bar{V}_x \bar{V}_p} \), where \( \bar{V}_x \) is the upper bound of \( V_x \), given by:

\[
\bar{V}_x = p_{\text{dis}}(a_{\text{dis}}^\text{med}) (a_{\text{min}} - \bar{a})^2 + p_{\text{dis}}(a_{\text{dis}}^\text{max}) (a_{\text{max}} - \bar{a})^2 + \sum_{i=1}^{i_{\text{max}}-1} p_{\text{dis}}(a_{\text{dis}}^i) (a_{\text{dis}}^i - \bar{a} - \frac{1}{2} \Delta)^2 + \sum_{i=1}^{i_{\text{min}}+1} p_{\text{dis}}(a_{\text{dis}}^i) (a_{\text{dis}}^i - \bar{a} + \frac{1}{2} \Delta)^2, \tag{6}
\]

where \( a_{\text{dis}}^i \) is the median of ith Alice’s discrete interval, \( p_{\text{dis}}(a_{\text{dis}}^i) \) is the probability of \( a_{\text{dis}}^i \), \( \bar{a} \) is the mean of Alice’s measurement result, \( a_{\text{min}} \) and \( a_{\text{max}} \) are the minimal
and maximum values of a large bound \([-a_{\text{lim}}, a_{\text{lim}}]\). This bound is set to make the probability that measurement result \(a\) is outside the bound is negligible for more rigorous proof. As shown in Eq. 6, we use the boundary of the discrete interval which is farther from 0 point to estimate the upper bound of \(V_x\), i.e., we use \(a_i^{\text{dis}} + \frac{1}{2} \Delta\) if \(a_i^{\text{dis}} > 0\) and use \(a_i^{\text{dis}} - \frac{1}{2} \Delta\) if \(a_i^{\text{dis}} < 0\). We give the same definitions and assumptions for \(\bar{V}_p\).

Our protocol uses vacuum state as randomness source which is not be disturbed by Eve. And we define the variance of vacuum fluctuation as \(\sigma^2_{\text{vac}} = 1\) for normalized calculation. In practice, we must consider the excess noise \(\epsilon\) which is mixed in quantum state \(\rho_A\) due to the side information. Thus, the expected variance of Alice’s measurement result becomes \(\sigma^2 = V_x = V_p = \epsilon + \sigma^2_{\text{vac}}\). Alice’s homodyne measurement result \(a\) is a continuous variable which satisfies Gaussian distribution with variance \(\sigma^2\) and zero mean. The expected discrete probability \(p_{\text{dis}}(a_i^{\text{dis}})\) of \(a_i^{\text{dis}}\) can be calculated out based on the Gaussian distribution.

No assumptions about the source have been made in the security analysis, if we do not distinguish between. Therefore, the final extracted randomness is source device independent and does not relate to Eve.

### 3.2 Finite-size effect

In fact, our CV-SI-QRNG and data sampling cannot run forever. Thus, we could only extract finite-size sequences for random numbers. Some parameters of the extracted data will have fluctuations and deviations compared with their ideal values. In our protocol, the number of total measurement results \(n_{\text{tot}}\) and the number of check samples \(n_c\) are finite absolutely. Therefore, the extractable randomness in Eq. 4 should be bounded with finite-size effect.

Our analysis of finite-size effect focuses on \(\hat{V}_x\) which is defined in Eq. 6 in infinite-size scenario. The formula we used for obtaining \(V_x\) with discrete sequence \(a = \{a_1, a_2 \cdots a_m\}\) which is the output of convertor is

\[
\hat{V}_x = \frac{1}{m} \sum_{i=1}^{m} \left( a_i \pm \frac{1}{2} \Delta - \bar{a} \right)^2,
\]

where \(m\) is the check data length and \(\hat{V}_x\) is the experimental value of \(V_x\) which is calculated from samples. For the plus or minus sign in front of \(\frac{1}{2} \Delta\), we take plus sign if \(a_i > 0\) and take minus sign if \(a_i < 0\). Now, we expand Eq. 7 to

\[
\hat{V}_x = \frac{1}{m} \sum_{i=1}^{m} (a_i^2 \pm \Delta \cdot a_i) + \frac{1}{m} \sum_{i=1}^{m} (\mp \Delta \cdot \bar{a}) + \frac{1}{4} \Delta^2.
\]

An approximation could be done that \(\pm \Delta \cdot a_i = |\Delta \cdot a_i|\) based on the selection of plus or minus sign. For a large \(m\), the numbers of plus sign and minus sign we
take in Eq. 7 are unlikely to differ by orders of magnitude, and \( \bar{a} \) is almost close to 0. Therefore, we can make an approximation that \( \frac{1}{m} \sum_{i=1}^{m} (\mp \Delta \cdot \bar{a}) \approx 0 \). Taking these results into Eq. 8, we get

\[
\hat{V}_x = \frac{1}{m} \sum_{i=1}^{m} \left( a_i^2 + |\Delta \cdot a_i| \right) - \bar{a}^2 + \frac{1}{4} \Delta^2 \tag{9}
\]

It should be taken note that \( \bar{a} \) has correlation with \( a_i \), but for a large \( m \) the correlation is very small, and \( \bar{a} \) is close to 0 compared with \( a_i^2 \). So we use ideal value \( \mu_a \) instead of \( \bar{a} \). We define \( a_i^2 + |\Delta \cdot a_i| = b_i \) and get

\[
\hat{V}_x = \frac{1}{m} \sum_{i=1}^{m} b_i - \mu_a^2 + \frac{1}{4} \Delta^2 \tag{10}
\]

Let’s focus on \( \sum_{i=1}^{m} b_i \). In Sect. 2, we have assumed that the experiments are repeated in i. i. d. way; thus, the extracted numbers \( a_i \) are i. i. d. too. For this reason, \( b_i \) is i. i. d. absolutely based on its definition. According to central limit theorem [38], for an i. i. d. random number sequence \( b = \{b_1, b_2 \cdots b_m\} \), if \( m \) is large enough (more than \( 10^4 \)), it has the following approximation:

\[
\frac{\sum_{i=1}^{m} b_i - m \mu_b}{\sqrt{m} \sigma_b} \sim N (0, 1) \tag{11}
\]

where \( \mu_b \) and \( \sigma_b \) are the ideal mean value and standard variance of \( b \), respectively, and \( N(0, 1) \) is standard Gaussian distribution. Thus, the value of \( \frac{1}{m} \sum_{i=1}^{m} b_i \) satisfies Gaussian distribution approximately:

\[
\frac{1}{m} \sum_{i=1}^{m} b_i \sim N \left( \mu_b, \frac{\sigma_b^2}{m} \right) \tag{12}
\]

and \( \hat{V}_x \) satisfies

\[
\hat{V}_x \sim N \left( \mu_b - \mu_a^2 + \frac{1}{4} \Delta^2, \frac{\sigma_b^2}{m} \right) \tag{13}
\]

In Eq. 6, an assumption has been made that the minimal and maximum values of \( a \) are determined by a large bound \([-a_{\text{lim}}, a_{\text{lim}}]\). So, we make the same assumption here for calibrating \( \mu_b, \sigma_b \) and \( \mu_a \). From the above derivation, we know that \( \mu_b - \mu_a^2 + \frac{1}{4} \Delta^2 \) is the approximation of ideal value of \( \hat{V}_x \), that is \( \bar{V}_x \) based on the definition of \( \hat{V}_x \) in Eq. 7. The gap between \( \mu_b - \mu_a^2 + \frac{1}{4} \Delta^2 \) and \( \bar{V}_x \) is 5.441 \times 10^{-6} \) with sampling range size of 3\( \sigma \), when check data length \( m = 10^6 \), sampling resolution \( n = 16 \), and confidence probability \( \epsilon = 10^{-10} \). This gap shortly reduces to 1.927 \times 10^{-9} \) as sampling range size increase to 5\( \sigma \), much less than the normal value of \( \bar{V}_x \). On account of this negligible gap, We can use \( \mu_b - \mu_a^2 + \frac{1}{4} \Delta^2 \) to substitute \( \bar{V}_x \) securely.
Now we can estimate the confidence interval of $\hat{V}_x$ with confidence probability $\epsilon$:

$$\hat{V}_x \in [\bar{V}_x - \Delta V, \bar{V}_x + \Delta V],$$  \hspace{1cm} (14)

where

$$\Delta V = Z_{\epsilon/2} \sqrt{\frac{\sigma_{b}^2}{m}},$$  \hspace{1cm} (15)

and $Z_{\epsilon/2}$ satisfies $1 - erf \left( \frac{Z_{\epsilon/2}}{\sqrt{2}} \right) = \epsilon$. $\epsilon$ is the probability of the $\hat{V}_x$ estimation outside the confidence interval. And $erf$ is the error function defined as

$$erf \ (x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} dt.$$  \hspace{1cm} (16)

To get the upper bound of $\lambda$ in Eq. 5, we use $\hat{V}_x$ to estimate $V_x$. So there is

$$V_{x \max} \approx \bar{V}_x + Z_{\epsilon/2} \sqrt{\frac{\sigma_{b}^2}{m}}.$$  \hspace{1cm} (17)

In practice, we use the variance of $a$, which is calculated from samples, as ideal variance $V_x$. We can obtain the ideal distribution of $a_i$, which is a Gaussian distribution with variance $V_x$. Then, we use the standard variance $\sigma_b$ and mean $\mu_b$ of $b_i$, which is calculated depend on the ideal distribution of $a_j$, as the ideal values. Even in some worse conditions that this ideal variance deviates too much from the real variance of our system, our protocol still keeps the system security by estimate the upper bound of $V_x$ with a given confidence probability. This is the significance of our work.

Similarly, we consider finite-size effect on $V_p$ with the same method, and get $V_{p \max}$. Putting these parameters into Eq. 1, we will get the upper bound of $\lambda$:

$$\lambda_{\max} = \sqrt{V_{x \max} V_{p \max}}.$$  \hspace{1cm} (18)

In fact, finding the confidence interval of $\lambda$ with confidence probability $\epsilon$ is a more precise method to estimate the upper bound of $\lambda$. But the calculation is too complex for the distribution of product of two Gaussian distributions which do not satisfy standard Gaussian distribution [39]. Fortunately, the value of $\lambda$ calculated by the upper bounds of $V_x$ and $V_p$ is very close to the more precise one [29]. So, we can use $\lambda_{\max}$ to estimate the extractable randomness.

### 4 Simulation results and discussion

In this section, we will have some numerical simulations and discuss the results for considering finite size on CV-SI-QRNG with the ideal measurement assumption. We assume that the input state is vacuum state here, i.e., the X and P quadratures satisfy the same probability distribution. It leads to $V_{x \max} = V_{p \max}$. The large bound $[-a_{\lim}, a_{\lim}]$,
Fig. 2 The impact of check data length $m$ for finite-size effect on extractable randomness $R_{\text{dis}}(a_i : E)$. We set the excess noise $\varepsilon = 0.1$, ideal variance $\sigma^2 = 1 + \varepsilon = 1.1$, confidence probability $\varepsilon = 10^{-10}$, sampling resolution $n = 16$, and sampling range size $N = 3\sigma$ here.

which is defined to determine the $a_{\text{max}}$ and $a_{\text{min}}$ in Eq. 6, is set as $[-10\sigma, 10\sigma]$. The probability that $a_i$ falls out of the large bound is less than $1.5 \times 10^{-23}$. Noted that, if we set check data length $m \to \infty$, that is in infinite size condition, the performance of our protocol will be identical to that in Ref. [23]. We will also compare the finite size condition with infinite size condition in the following discus for highlighting the impact of finite size effect.
Fig. 4 The impact of sampling range size $N$ for finite-size effect on extractable randomness $R_{\text{dis}}(a_i : E)$ with different check data length $m$, which from bottom to top curve correspond to $m = 10^4, 10^5, 10^6, 10^7, 10^8$. We set the excess noise $\epsilon = 0.1$, ideal variance $\sigma^2 = 1 + \epsilon = 1.1$, sampling resolution $n = 16$, and confidence probability $\epsilon = 10^{-10}$ here.

Fig. 5 The impact of sampling range size $N$ for finite-size effect on extractable randomness $R_{\text{dis}}(a_i : E)$ with different sampling resolution $n$, which we choose $n = 8, 12, 16$. We set the excess noise $\epsilon = 0.1$, ideal variance $\sigma^2 = 1 + \epsilon = 1.1$, check data length $m = 10^4$, and confidence probability $\epsilon = 10^{-10}$ here.

Figure 2 shows the relationship between extractable randomness and check data length. The red solid line is the ideal extractable randomness in infinite-size case, and the blue dot–dashed line is the extractable randomness with the effect of finite size. We can see that, the blue line changes sharply since check data length reduces under $10^4$ (some assumptions we made above have large deviation here), i.e., finite-size effect has remarkable influence on finial randomness. When check data length is larger than

Springer
The finite-size analysis of CV-SI-QRNG

Fig. 6 Simulation results for upper bound of Eve’s information $S(a_i; E)$ and extractable randomness $R_{\text{dis}}(a_i; E)$ as a function of sampling resolution $n$ under ideal infinite-size and finite-size condition. We set the excess noise $\epsilon = 0.1$, ideal variance $\sigma^2 = 1 + \epsilon = 1.1$, sampling range size $N = 3\sigma$, check data length $m = 10^4$, and confidence probability $\epsilon = 10^{-10}$ here.

10^7, the finite-size line is very close to ideal line. In practical, we should choose an appropriate check data length to obtain enough final randomness and meanwhile use less computing resources.

The another important parameter of finite-size effect is confidence probability $\epsilon$. The relationship between extractable randomness and confidence probability is shown in Fig. 3. The curve of finite size is more smoothly as $\epsilon$ decreases. At the point $\epsilon = 1$, the finite-size effect on extractable randomness disappears due to the estimation for confidence interval of variance converges toward its ideal value. The confidence probability can be used as security parameter to adjust final randomness for different demands on security.

The function of sampling range size and extractable randomness under finite-size effect with different check data lengths is displayed in Fig. 4. The curve is closer to the ideal curve (black solid line) with longer check data length. Finite-size effect is relatively smooth after the increase in sampling range size over the value that achieves max extractable randomness. In fact, as the increase in sampling range size, deviation between ideal and finite size reduces very slowly. The gap between them is 0.1943 bit when $N = 10\sigma$, and reduces to 0.1924 bit when $N = 100\sigma$ with check data length $m = 10^4$, sampling resolution $n = 16$, and confidence probability $\epsilon = 10^{-10}$. We would like to point out that the most serious impact of finite size appears in the peak value of extractable randomness, the same in Fig. 5, due to the comparatively uniform distribution of $a_i$. We discuss it in details in Appendix A.

In Fig. 5, we show the function of sampling range size and extractable randomness under finite-size effect with different sampling resolution. We choose $n = 8, 12, 16$ three common sampling resolutions to show the influence. And we can see that, the finite-size effect has similar impact with different sampling resolutions. Figure 6 shows...
Fig. 7 The probability distribution $P_{\text{dis}}(a_i)$ of Alice’s quantified output with different sampling range size $N = 2\sigma, 4\sigma, 8\sigma, 100\sigma$. We set the excess noise $\epsilon = 0.1$, ideal variance $\sigma^2 = 1 + \epsilon = 1.1$ and fixed sampling resolution $n = 8$ here.

this impact in detail. We fix sampling range to $N = 3\sigma$ and display the extractable randomness $R_{\text{dis}}(a_i : E)$ and the upper bound of Eve’s information $S(a_i : E)$ as a function with sampling resolution $n$ under ideal infinite-size and finite-size conditions. The gap between the ideal curve and finite-size one is almost fixed with different sampling resolutions for both $R_{\text{dis}}(a_i : E)$ and $S(a_i : E)$.

5 Conclusion

In this paper, we propose a method considering impact of finite-size effect on continuous variable source-independent quantum random number generation and improve the security of our protocol in practical applications. The method provides some formulas through central limit theorem for calculating the statistical fluctuations of parameters under finite-size scenario. Some numerical simulations are provided to show the influence on extractable randomness with several key variable parameters. We discuss these results in detail and summarize some conclusions, which could be useful in practical realization.

The outcomes we found in numerical simulations show the peculiarity of finite-size effect. The final randomness is closer to the ideal value, with the increase in check data length and confidence probability, i.e., the effect of finite size is monotonically related with these two parameters mightily. One could choose appropriate values according to the need for security and available computing resources in practical. An apposite sampling range size will reach the maximum of extractable randomness, which makes
a relatively uniform distribution of input and the largest loss of randomness due to finite-size effect. Moreover, the loss is smoothly decreasing after sampling range size rises over the peak, while the impact of sampling resolution is more steady.

Our method still needs some improvements which can get further study, such as the treatment of the most border sampling value and the approximate on $\lambda_{\text{max}}$. Noted that, SI-QRNG is named for our protocol to unify the previous work [20,22,23,25], which indicates the input quantum state can be arbitrary quantum state, such as vacuum state, thermal state, squeezed state and so on. Namely, the “no assumption about source” we mentioned above means we do not assume a specific input quantum state. The measurement in our protocol is assumed to be repeated in an independently and identically distributed (i. i. d.) way to ensure the exacted random numbers is i. i. d, where the notation of semi-source independent will be more precise. In fact, the i. i. d. input is the optimal choice for extracting i. i. d. randomness efficiently [40]. In addition, this work can promote the applications of continuous variable source-independent quantum random number generation and ensure the security in its processing.

Acknowledgements This work was supported by the Key Program of National Natural Science Foundation of China under Grant No. 61531003 the National Natural Science Foundation under Grant No. 62001044, the Fund of CETC under Grant No. 6141B08231115, and the Fund of State Key Laboratory of Information Photonics and Optical Communications.

Appendix: Effect of sampling range size

In this appendix, we focus on the impact of sampling range size on finite-size effect by analyzing the loss of extractable randomness, because of the remarkable influence of it for extractable randomness in both ideal and finite-size conditions.

This parameter mainly influences the probability distribution of Alice’s discrete results as shown in Fig. 7. The probability will concentrate in the middle bins with the increase in the sampling range size. And if the sampling range size $N$ is too small, most of the probability will lie in boundary bins.

Figure 8 shows the extractable randomness $R_{\text{dis}}(a_i : E)$ with a larger sampling range size bound compared with Fig. 4, and the difference value between ideal and finite-size conditions is shown in the inner mini-figure. We can see that the curve of difference values has a peak when $N = 2.7\sigma$, and extractable randomness rises and falls sharply near the peak. But after the sampling range size increases over $5\sigma$, the reduction of difference values suddenly becomes very smooth.

The sampling range size value $N$ near peak is exactly the value that makes the most uniform probability distribution of $a_i$. When the sampling range size is small, compared to that with large sampling range size, the change of probability distribution is more conspicuous as the sampling range size changes. This corresponds to the sharp and smooth changes in the mini-figure of Fig. 8, respectively.

It should be noted that the peak of extractable randomness on ideal infinite case appears in $N = 3.4\sigma$ and does not coincide with the peak of difference. It leads to the max extractable randomness under finite-size condition which appears in $N = 3.7\sigma$, 

\[ \text{Springer} \]
Fig. 8 Simulation result for extractable randomness $R_{dis}(a_i : E)$ as a function of sampling range size $N$ under ideal infinite-size and finite-size scenario. The mini-figure inside is the difference between these two scenarios. We set the excess noise $\varepsilon = 0.1$, ideal variance $\sigma^2 = 1 + \varepsilon = 1.1$, check data length $m = 10^4$, confidence probability $\epsilon = 10^{-10}$ and sampling resolution $n = 8$ here

slightly larger than the value in ideal case. It should be paid attention in practical for obtaining maximal final randomness.

References

1. Herrero-Collantes, M., Garcia-Escartin, J.C.: Quantum random number generators. Rev. Mod. Phys. 89(1), 015004 (2017)
2. Ma, X., Yuan, X., Cao, Z., Qi, B., Zhang, Z.: Quantum random number generation npj. Quantum Inf. 2, 16021 (2016)
3. Nguyen, L., Rehain, P., Sua, Y.M., Huang, Y.: Programmable quantum random number generator without postprocessing. Opt. Lett. 43(4), 631–634 (2018)
4. Wang, J., Xie, T., Zhang, H., Yang, D., Xie, C., Wang, J.: A bias-free quantum random number generation using photon arrival time selectively. IEEE Photonics J. 7(2), 1–8 (2015)
5. Yan, Q., Zhao, B., Hua, Z., Liao, Q., Yang, H.: High-speed quantum-random number generation by continuous measurement of arrival time of photons. Rev. Sci. Instrum. 86(7), 073113 (2015)
6. Stefanov, A., Gisin, N., Guinnard, O., Guinnard, L., Zbinden, H.: Optical quantum random number generator. J. Mod. Opt. 47, 595–598 (2000)
7. Wayne, M.A., Migdall, A.L., Levine, Z.H., Bienfang, J.C.: A post-processing-free single-photon random number generator with ultra-low latency. Opt. Express 26(25), 32788–32801 (2018)
8. Ren, M., Wu, E., Liang, Y., Jian, Y., Guang, W., Zeng, H.: Quantum random-number generator based on a photon-number-resolving detector. Phys. Rev. A 83, 023820 (2011)
9. Zheng, Z., Zhang, Y., Huang, W., Yu, S., Guo, H.: 6 Gbps real-time optical quantum random number generator based on vacuum fluctuation. Rev. Sci. Instrum. 90(4), 043105 (2019)
10. Gehring, T., Lupo, C., Kordts, A., Nikolic, D.S., Jain, N., Pedersen, T.B., Pirandola, S., Andersen, U. L. 8 GBit/s real-time quantum random number generator with non-iid samples. arXiv preprint arXiv:1812.05377 (2018)
11. Shakhovoy, R., Sych, D., Sharoglazova, V., Udaltsov, A., Fedorov, A., Kurochkin, Y.: Quantum noise extraction from the interference of laser pulses in optical quantum random number generator. arXiv preprint arXiv:1910.00219 (2019)
12. Wei, W., Guo, H.: Bias-free true random-number generator. Opt. Lett. 34, 1876 (2009)
13. Huang, M., Chen, Z., Zhang, Y., Guo, H.: A Gaussian-distributed quantum random number generator using vacuum shot noise. Entropy 22(6), 618 (2020)
14. Huang, M., Chen, Z., Zhang, Y., Guo, H.: A phase fluctuation based practical quantum random number generator scheme with delay-free structure. Appl. Sci. 10(7), 2431 (2020)
15. Huang, W., Zhang, Y., Zheng, Z., Li, Y., Xu, B., Yu, S.: Practical security analysis of a continuous-variable quantum random-number generator with a noisy local oscillator. Phys. Rev. A 102(1), 012422 (2020)
16. Pironio, S., Acín, A., Massar, S., de La Giroday, A.B., Matsukevich, D.N., Maunz, P., Olmschenk, S., Hayes, D., Luo, L., Manning, T.A., Monroe, C.: Random numbers certified by Bell’s theorem. Nature 464(7291), 1021 (2010)
17. Liu, Y., Yuan, X., Li, M., Zhang, W., Zhao, Q., Zhong, J., Cao, Y., Li, Y., Chen, L., Li, H., Peng, T., Chen, Y., Peng, C., Shi, S., Wang, Z., You, L., Ma, X., Fan, J., Zhang, Q., Pan, J.: High-speed device-independent quantum random number generation without a detection loophole. Phys. Rev. Lett. 120(1), 010503 (2018)
18. Nie, Y., Guan, J., Zhou, H., Zhang, Q., Ma, X., Zhang, J., Pan, J.W.: Experimental measurement-device-independent quantum random number generation. Phys. Rev. A 94, 060301 (2016)
19. Cao, Z., Zhou, H., Ma, X.: Loss-tolerant measurement-device-independent quantum random number generation. New J. Phys. 17, 125011 (2015)
20. Cao, Z., Zhou, H., Yuan, X., Ma, X.: Source-independent quantum random number generation. Phys. Rev. X 6(1), 011020 (2016)
21. Smith, P.R., Marangon, D.G., Lucamarini, M., Yuan, Z.L., Shields, A.J.: Simple source-device-independent continuous-variable quantum random number generator. Phys. Rev. A 99(6), 062326 (2019)
22. Marangon, D.G., Vallone, G., Villoresi, P.: Source-device-independent ultrafast quantum random number generation. Phys. Rev. Lett. 118(6), 060503 (2017)
23. Xu, B., Chen, Z., Li, Z., Yang, J., Qi, S., Huang, W., Zhang, Y., Guo, H.: High speed continuous variable source-independent quantum random number generation. Quantum Sci. Technol. 4(2), 025013 (2019)
24. Avesani, M., Marangon, D.G., Vallone, G., Villoresi, P.: Source-device-independent heterodyne-based quantum random number generator at 17 Gbps. Nat. Commun. 9(1), 5365 (2018)
25. Zheng, Z., Zhang, Y., Huang, M., Chen, Z., Yu, S., Guo, H.: Bias-free source-independent quantum random number generator. Opt. Express 28(15), 22388–22398 (2020)
26. Vallone, G., Marangon, D.G., Tomasin, M., Villoresi, P.: Quantum randomness certified by the uncertainty principle. Phys. Rev. A 90(5), 052327 (2014)
27. Wang, X., Zhang, Y., Yu, S., Guo, H.: High-speed implementation of length-compatible privacy amplification in continuous-variable quantum key distribution. IEEE Photonics J. 10(3), 1–9 (2018)
28. Li, Q., Yan, B.-Z., Mao, H.-K., Xue, X.-F., Han, Q., Guo, H.: High-speed and adaptive fpga-based privacy amplification in quantum key distribution. IEEE Access 7, 21482–21490 (2019)
29. Leverrier, A., Grosshans, F., Grangier, P.: Finite-size analysis of a continuous-variable quantum key distribution. Phys. Rev. A 81(6), 062343 (2010)
30. Jouguet, P., Kunz-Jacques, S., Diamanti, E., Leverrier, A.: Analysis of imperfections in practical continuous-variable quantum key distribution. Phys. Rev. A 86(3), 032309 (2012)
31. Zhang, X., Zhang, Y., Zhao, Y., Wang, X., Yu, S., Guo, H.: Finite-size analysis of continuous-variable measurement-device-independent quantum key distribution. Phys. Rev. A 96(4), 042334 (2017)
32. Devetak, I., Winter, A.: Distillation of secret key and entanglement from quantum states. P. Roy. Soc. A-Math. Phy. 461(2053), 207–235 (2005)
33. Zhang, Y., Chen, Z., Pirandola, S., Wang, X., Zhou, C., Chu, B., Zhao, Y., Xu, B., Yu, S., Guo, H.: Long-distance continuous-variable quantum key distribution over 202.81 km of fiber. Phys. Rev. Lett. 125(1), 010502 (2020)
34. Zhang, Y., Li, Z., Chen, Z., Weedbrook, C., Zhao, Y., Wang, X., Huang, Y., Xu, C., Zhang, X., Wang, Z., et al.: Continuous-variable QKD over 50 km commercial fiber. Quantum Sci. Technol. 4(3), 035006 (2019)
35. Zhang, X., Zhang, Y., Li, Z., Yu, S., Guo, H.: 1.2-GHz balanced homodyne detector for continuous-variable quantum information technology. IEEE Photonics J. 10(5), 1–10 (2018)
36. Wolf, M.M., Giedke, G., Cirac, J.I.: Extremality of Gaussian quantum states. Phys. Rev. Lett. 96(8), 080502 (2006)
37. Ma, X., Xu, F., Xu, H., Tan, X., Qi, B., Lo, H.-K.: Postprocessing for quantum random-number generators: entropy evaluation and randomness extraction. Phys. Rev. A 87(6), 062327 (2013)
38. Rosenblatt, M.: A central limit theorem and a strong mixing condition. P. Natl. Acad. Sci. USA 42(1), 43 (1956)
39. Springer, M.D., Thompson, W.E.: The distribution of products of beta, gamma and Gaussian random variables. SIAM J. Appl. Math. 18(4), 721–737 (1970)
40. Drahi, D., Walk, N., Hoban, M.J., Kolthammer, W.S., Nunn, J., Barrett, J., Walmsley, I.A.: Certified quantum randomness from untrusted light. arXiv preprint arXiv:1905.09665 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.