Experimental fast quantum random number generation using high-dimensional entanglement with entropy monitoring

Citation: Xu, Feihu et al. "Experimental Fast Quantum Random Number Generation Using High-Dimensional Entanglement with Entropy Monitoring." Optica 3, 11 (October 2016): 1266-1269. © 2016 Optical Society of America

As Published: http://dx.doi.org/10.1364/optica.3.001266

Publisher: Optical Society of America

Persistent URL: http://hdl.handle.net/1721.1/111034

Version: Original manuscript: author's manuscript prior to formal peer review

Experimental fast quantum random number generation using high-dimensional entanglement with entropy monitoring

Feihu Xu,[1,*] Jeffrey H. Shapiro,[1] and Franco N. C. Wong[1]

[1]*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

A quantum random number generator (QRNG) generates genuine randomness from the intrinsic probabilistic nature of quantum mechanics. The central problems for most QRNGs are estimating the entropy of the genuine randomness and producing such randomness at high rates. Here we propose and demonstrate a proof-of-concept QRNG that operates at a high rate of 24 Mbit/s by means of a high-dimensional entanglement system, in which the user monitors the entropy in real time via the observation of a nonlocal quantum interference, without a detailed characterization of the devices. Our work provides an important approach to a robust QRNG with trusted but error-prone devices.

I. INTRODUCTION

Randomness is indispensable for a wide range of applications, ranging from Monte Carlo simulations to cryptography. Quantum random number generators (QRNGs) can generate true randomness by exploiting the fundamental indeterminism of quantum mechanics [1]. Most current QRNGs are photonic systems built with trusted and calibrated devices [2–6] that provide Gbit/s generation speeds at relatively low cost. However, a central issue for these QRNGs is how to certify and quantify the entropy of the genuine randomness, i.e., the randomness that originates from the intrinsic unpredictability of quantum-mechanical measurements. Entropy estimates for specific setups were recently proposed using sophisticated theoretical models [8–10]. Nevertheless, these techniques require complicated device characterization that may be difficult to accurately assess in practice.

A solution to estimating the entropy is the device-independent (DI) or self-testing QRNG [11–13], but its practical implementation is challenging because it requires loophole-free violation of Bell's inequality, resulting in low generation rates of $\sim$1 bit/s [12, 13]. Recently, Lunghi *et al.* proposed a more practical solution that is based on a dimension witness [14], in which the randomness can be guaranteed based on a few general assumptions that do not require detailed device characterization. This scheme is highly desirable as it focuses on real-world implementations with trusted but error-prone devices, although its implementation with a two-dimensional (qubit) system still suffers from low generation rates of 10's of bits/s [14].

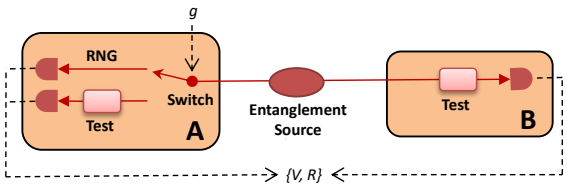
In this paper, we propose and experimentally demonstrate a fast QRNG operating at a rate of 24 Mbits/s, in which we can quantify and monitor, in real time, the entropy of genuine randomness without a detailed characterization of the trusted but error-prone devices. Our approach uses time-energy entangled photon pairs with high-dimensional temporal encoding [15]. High-dimensional temporal encoding is advantageous when the average interval between photon detection events is much longer than a time-bin duration set by the detector timing resolution. Such situations arise in typical quantum information processing tasks when single-photon detectors have long recovery times or when the pair generation rate must be kept low to minimize multi-pair events [13, 14]. The amount of genuine quantum randomness is quantified and monitored directly from observation of a nonlocal interference [16], and it is separated from other sources of randomness such as technical noise with a randomness extractor. We achieved the high generation rate by virtue of three experimental features: a high-dimensional time-energy entangled-photon source capable of producing multiple random bits per photon, a high-visibility Franson interferometer for evaluating entanglement, and high-efficiency superconducting nanowire single-photon detectors (SNSPDs). As a consequence, we are able to demonstrate a high-performance QRNG with a tolerance of device imperfections.

II. PROTOCOL

Table I summarizes our protocol. An entanglement source generates high-dimensional entangled photon pairs. As an example, in the ideal case, the N_d-dimensional biphoton entangled state can be written as $|\psi\rangle = \frac{1}{\sqrt{N_d}} \sum_{i=0}^{N_d-1} |i\rangle_{\rm A} \bigotimes |i\rangle_{\rm B}$, where $|i\rangle$ represents a single photon at a discretized time interval i. This state is observed by two measurement systems, one for random number generation (RNG) and the other for testing. Randomness is generated in the RNG mode from the state held by system A, $\rho_{\rm A}$, which we assume is not pure and is correlated with environmental noise that models device imperfections. In the testing mode, the joint state is measured.

A key assumption of our approach is that the devices in the protocol are *trusted*, namely they are not deliberately designed to fool the user, but the implementation may be imperfect. The central task is to estimate and monitor the amount of genuine randomness based *only* on measurements. This is a nontrivial task as the observed randomness can have different origins. If the state $\rho_{\rm A}$ is a superposition of high-dimensional states, then the outcome R_i cannot be predicted with certainty, even if the internal state is known, thus resulting in genuine quantum randomness. On the other hand, the randomness may be due to technical imperfections such as detector noise and temperature fluctuations, whose randomness clearly has *no* quantum origin, since the outcome R_i can be perfectly guessed if the imperfections were well quantified.

*Electronic address: fhxu@mit.edu



1. There are N measurement rounds. The entanglement source generates a high-dimensional entangled photon pair in each round. Bits $g_1, g_2, \ldots, g_N$ with values 0 or 1 are independently chosen at random according to a $(q, 1-q)$ distribution.
2. For each round $i \in [N]$, if $g_i = 0$, it is a generation round, then device **A** performs a time measurement in the 'RNG' basis and outputs a random number R_i. If $g_i = 1$, it is a testing round, then devices **A** and **B** perform a joint frequency measurement in the 'Test' basis.
3. From the results in the 'Test' basis, calculate the testing value V. If V exceeds a pre-set value V_0, then the protocol **succeeds**. Otherwise, it **aborts**.
4. If the protocol succeeds, the randomness throughput is evaluated from V and a randomness extraction is applied to $\{R_i\}$ to produce the genuine random numbers.

TABLE I: Protocol for quantum random number generation.

In our approach, the amount of genuine randomness is monitored from the Franson visibility V [16]. In particular, classical fields result in V that is no greater than 50%. For a maximally entangled state, V would be 100% in the ideal case [19]. Conceptually, $V > 50\%$ guarantees that the source's output is entangled and thus contains genuine randomness. Rigorously, Ref. [20] has proven that, if we assume the biphoton wave function is Gaussian, V provides an explicit bound for the correlations in frequency measurement, which in turn upper-bounds the conditional maximum entropy (given system **B**) via the theory developed in [17]. By using the entropic uncertainty relation for smooth entropies [18], we can determine the conditional min-entropy given the environmental noise and thus the guessing probability, i.e., the amount of genuine randomness (see Appendix A).

Our approach is different from Ref. [14] that is based on a dimension witness and was restricted to a two-dimensional system. Our system allows a much higher dimensionality that can be chosen in the post-processing step [15], which was $N_d = 2048$ dimensions in our experiment (see below). Our protocol provides self-monitoring because measurements of V directly quantify the amount of genuine randomness in the observed data. A threshold value V_0 is pre-selected and the randomness can be generated *only* when the observation satisfies $V > V_0$. Two particular advantages of this approach are: (i) the observation of V does not rely on detailed models of the devices that are employed; and (ii) no loophole-free Bell inequality violation is required.

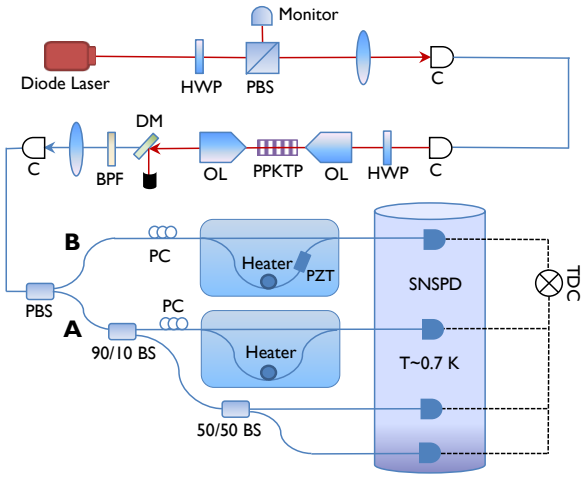


FIG. 1: Experimental setup. A cw diode laser pumps a PPKTP waveguide to generate time-energy entangled photon pairs. The orthogonally polarized signal and idler photons are coupled into a single-mode fiber, separated, and directed to **A** and **B**, respectively. The signal photons in **A** are passively selected by a 90/10-ratio beam splitter for time-of-arrival measurement to generate random numbers or Franson measurement to check entanglement quality. HWP: half wave plate; PBS: polarization beam splitter; C: coupler; OL: objective lens; DM: dichroic mirror; BPF: band-pass filter; PC: polarization controller; BS: beam splitter; PZT: piezoelectric transducer; SNSPD: superconducting nanowire single-photon detector; TDC: time-to-digital converter.

III. EXPERIMENT

Figure 1 shows the experimental setup. Time-energy entangled photon pairs were generated via spontaneous parametric down-conversion (SPDC) in a periodically-poled $KTiOPO_4$ (PPKTP) waveguide [21] that outputs multiple spatial modes at telecom wavelengths. The 46.1 μm grating period was designed for type-II quasi-phase-matched wavelength-degenerate outputs at 1560 nm in the fundamental modes of the signal and idler fields. The phase-matching bandwidth was 1.6 nm with a corresponding biphoton correlation time of 2 ps. The pump was a 780-nm continuous-wave (cw) diode laser with a measured coherence time of 2.2 μs and the pump power coupled into the waveguide was monitored by a power meter. We extracted the fundamental signal and idler modes using a dichroic mirror to remove the pump and a 10-nm band-pass filter to spectrally remove the higher-order SPDC spatial modes. The fundamental modes were coupled into a standard single-mode fiber, achieving a $\sim$81% waveguide-to-fiber coupling efficiency. We used a polarization beam splitter to separate the orthogonally polarized signal and idler photons and send them to devices **A** and **B**, respectively. Losses in the waveguide and from the waveguide to the fiber were $\sim$15% and $\sim$12%, respectively. Overall, we measured a system efficiency of $\sim$50% including the single-photon detector efficiency.

A Franson interferometer is ideally suited for measuring the entanglement quality of a cw-pumped source of time-energy entangled light [15, 20]. We set up a Franson interferometer with local dispersion cancellation that was comprised of two identical unbalanced Mach-Zehnder interferometers (MZIs), in which the long arm was made of standard single-mode fiber and low-dispersion LEAF fiber, such that the differential group delay (due to dispersion) between the long and short arms was zero [19]. To achieve long-term stability, the MZIs were enclosed in a multilayered thermally insulated box, whose temperature was actively stabilized. The long-short path mismatch of each MZI was measured to be $\Delta T = 3.2$ ns. We coiled the long-path fiber of each MZI on a closed-loop temperature-controlled heater to precisely match the ΔT of the two MZIs. The variable relative phase shift between the two MZIs was set by a piezoelectric transducer fiber stretcher. By carefully fine-tuning the input polarizations and the temperatures, our time-energy entanglement source was found to have a Franson interference visibility V of $98.8 \pm 0.3\%$, as shown in Fig. 2.

We performed a proof-of-concept implementation for the random basis choice, passively with a 90/10 beam splitter, i.e., $q = 0.9$ in Table I. The photon arrival times were measured by WSi SNSPDs [22] that were placed in a closed-cycle cryogenic system with sub-Kelvin operating temperatures (see Fig. 1). The SNSPDs were measured to have detection efficiencies of $\sim$85%, dark-count rates of $\sim$400/s, timing jitters of $\sim$250 ps, and maximum count rates of $\sim$2 MHz without detector saturation. To mitigate the long reset times of the SNSPDs and to achieve a higher generation rate, system **A** used a passive 50/50 beam splitter to distribute incident photons equally between two WSi SNSPDs and their data were interleaved. Hence, a total of four WSi SNSPDs were used and their detection-time outputs were recorded by time-to-digital converters.

In the experiment, both the time-bin duration δ and the frame size (dimensionality) N_d were chosen in the data post-processing step by parsing the raw timing records into the desired symbol length. As long as the frame duration $N_d\delta$ is smaller than the pump coherence time, we can precisely characterize the dimensionality. For experimental simplicity, we set δ equal to the detector timing jitter. N_d was optimized to be 2048 in order to produce the maximal genuine randomness per photon (Appendix A).

In our proof-of-principle experiment, we recorded data for a maximum duration of 60 s. We monitored the Franson visibility before and after data recording to ensure that the experimental V exceeded the preset threshold V_0 in order to extract nonzero random bits from the data. We set $V_0 = 98.5\%$ because it was the lower bound in most of the measurement runs in our experiment. We note that the amount of genuine randomness increases monotonically with increasing V_0, as shown in Fig. 3, and therefore it is desirable to use high-quality devices in order to achieve high Franson visibility $V > V_0$.

We evaluated the auto-correlation of the raw generation-round data to be $\sim$0.001, which satisfies the independent, identically distributed (iid) assumption made in our analysis. Using the theory developed in Appendix A, we obtain 6.0 bits/photon genuine randomness at $N_d = 2048$. After considering the unpaired-to-paired ratio of the SPDC output that we measured to be 1.8% (see Appendix B), we extracted about 5.9 bits per $\log_2(2048)$-bit sample. We implemented a Toeplitz-hashing extractor [8] to extract genuine random numbers. A Toeplitz-hashing extractor extracts a random bit-string m by multiplying the raw sequence n with the Toeplitz matrix (n-by-m matrix, random seed). The seed length of random bits required to construct the Toeplitz matrix is $d = n+m-1$. In our implementation with Matlab on a standard desktop computer, we chose the input and output bit-string lengths to be $n = 4096$ and $m = 4096 \times 5.9/11 \geq 2196$. Hence, a 4096-by-2196 Toeplitz matrix was generated in constructing the Toeplitz-hashing extractor. The output random bits successfully passed all the tests in the DIEHARD test suite (see Appendix Table III).

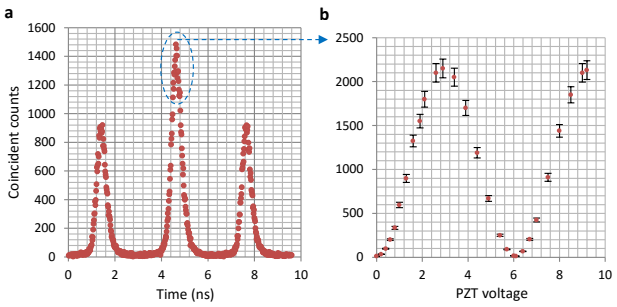


FIG. 2: Franson interference. (a) Typical distribution of signal and idler arrival-time difference in coincidence measurements. (b) Observed interference fringe at the central peak of (a) versus relative phase shift (proportional to PZT voltage). The observed raw visibility (without any subtraction) is $98.8 \pm 0.3\%$.

Figure 3 shows the experimental results for QRNG throughput for different running times. A longer running time produces more data, thus minimizing the finite-data effect and yielding more randomness from the raw data. The results show that a continuous running time of $\sim$60 s can already produce randomness that is close to the asymptotic case of infinitely long operation. The measured count rates of the two SNSPDs for signal photon arrival times were 1.8 and 2.3 Mcounts/s. Hence, the final QRNG rate is $5.9 \times 4.1 = 24.2$ Mbit/s, which is many orders of magnitude higher than previous experiments based on self-testing [12, 13] or a dimension witness [14]. This dramatically faster rate benefits from the following factors: high-dimensional entanglement system that generates *multiple* bits per photon, high-efficiency SNSPDs, high-quality PPKTP waveguide SPDC source with high fiber-coupling efficiency, and high-visibility Franson interferometry.

IV. CONCLUSION

To sum up, we have demonstrated a QRNG based on high-dimensional entanglement with a rate over 24 Mbit/s.

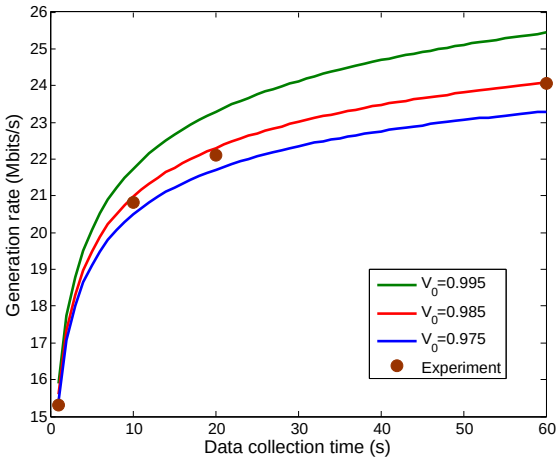


FIG. 3: QRNG throughput versus data collection running time. Solid red circles are experimental rates with $V_0 = 98.5\%$. The three curves are numerically evaluated results with different V_0 values, showing that a higher V_0 yields a higher throughput. For running times below 10 s, the finite-data effect reduces the QRNG throughput substantially, whereas for a running time over 50 seconds, the throughput is already close to its asymptotic limit of 24.2 Mbit/s at $V_0 = 98.5\%$.

Compared to the standard device-dependent approaches with fully calibrated devices, our QRNG delivers a stronger form of security requiring less characterization of the physical implementation. The performance is close to commercial QRNGs [23]. Though our approach offers a weaker form of security than self-testing QRNG, it focuses on a scenario with trusted but error-prone devices. Together with other types of QRNGs demonstrated very recently in [24, 25], we believe that our results constitute an important step towards generating truly random numbers for practical applications.

Our proof-of-principle experiment can be further improved in the following directions. First, the Franson visibility was mainly limited by temperature fluctuations. Integrated photonics can improve the temperature stability, thus leading to higher interference visibility, as demonstrated recently in [26]. Second, the system can be operated at a different wavelength such as the visible or near-IR region, where inexpensive high-efficiency Si single-photon detectors can be used to replace the SNSPDs and potentially allow the system to be integrated with silicon photonics technology. Third, the randomness extraction was processed off-line by software, which can be improved with a field-programmable gate array implementation for real-time extraction. Fourth, the monitoring of the visibility can be done in real time by continuously observing the Franson measurement. Lastly, to increase the security of the QRNG protocol, we note that a loophole-free Bell's inequality test has been proposed for time-energy entanglement [27], thus making it possible to extend our high-dimensional entanglement system to function as a self-testing QRNG. Our QRNG is just one example of high-dimensional quantum information processing that we believe is an important area for future study and practical applications.

Funding Information

Office of Naval Research (ONR) (N00014-13-1-0774); Air Force Office of Scientific Research (AFOSR) (FA9550-14-1-0052).

Acknowledgments

The authors thank Murphy Yuezhen Niu, Tian Zhong, Xiaodi Wu, Zheshen Zhang for many helpful discussions.

Appendix A: Quantification of genuine randomness

The amount of genuine randomness is quantified from a pair of incompatible quantum measurements, namely time measurement $\mathbb{T}$ and frequency measurement $\mathbb{W}$. We take $\mathbb{T}$ and $\mathbb{W}$ to be positive-operator valued measures (POVMs) on $\mathbf{A}$ with elements $\{\hat{\mathcal{M}}_t\}$ and $\{\hat{\mathcal{N}}_w\}$, and random outcomes T and W. Given an N_d-bin state observed by $\mathbf{A}$, the number of true random bits that can be extracted from T that are independent of the environment system $\mathbf{E}$ is given by the conditional min-entropy $H_{\min}(T|\mathbf{E})$. Specifically, the probability of guessing T by holding the system $\mathbf{E}$ is given by $p_{guess}(T|\mathbf{E}) = 2^{-H_{\min}(T|\mathbf{E})}$ [12]. Hence, $H_{\min}(T|\mathbf{E})$ quantifies the genuine randomness. We bound $H_{\min}(T|\mathbf{E})$ based on the uncertainty relation [17, 28], as proposed in [18]. While Ref. [18] demonstrated a QRNG with a maximal dimensionality of $N_d = 4$, our system is capable for a much higher dimensionality, i.e., $N_d > 2000$.

Consider three quantum systems $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{E}$ and the tripartite state ρ_{ABE}. The uncertainty relation can be written as [28]

$$H_{\min}(T|\mathbf{E}) \geq -\log_2 c - H_{\max}(W|\mathbf{B}), \tag{A1}$$

where $H_{\max}(W|\mathbf{B})$ denotes the maximum entropy for W given system $\mathbf{B}$, and c is the maximum "overlap" between the two POVMs [17, 28]. By assuming that $\hat{\mathcal{M}}_t$ and $\hat{\mathcal{N}}_w$ are projective measurements corresponding to mutually-unbiased N_d-dimensional bases, then $c = 1/N_d$. Based on [17], we bound the maximum entropy as follows:

$$H_{\max}(W|\mathbf{B}) \leq \log_2 \gamma(d_w^{\mathrm{L}_1} + \lambda), \tag{A2}$$

where $d_w^{\mathrm{L}_1}$ is the L_1 distance for correlations in the $\mathbb{W}$ basis,

$$\gamma(x) = (x + \sqrt{1+x^2})\Big(\frac{x}{\sqrt{1+x^2}-1}\Big)^x, \tag{A3}$$

and

$$\lambda \approx N_d \sqrt{\frac{1}{q(1-q)n_{\mathrm{T}}} \ln \frac{1}{\epsilon_1/4 - 2f(p_\alpha, n_{\mathrm{T}})}}, \tag{A4}$$

which quantifies the statistical fluctuations in the measurements. In Eq. (A4): n_{T} is the total number of detections in the $\mathbb{T}$ measurements; $f(p_\alpha, n_{\mathrm{T}}) = \sqrt{2(1-(1-p_\alpha)^{n_{\mathrm{T}}})}$; p_α

is the probability of a biphoton's signal photon arriving outside a frame; and ϵ_1 is the failure probability for the finite-data analysis, which was set to $\epsilon_1 = 10^{-10}$ in our experiment.

The L_1 distance $d_w^{L_1}$ can be bounded from the time-frequency uncertainty relation, as we now explain. The variances of the signal-idler time and frequency differences can be written as [20]

$$\begin{aligned} \langle(\Delta t)^2\rangle &= \frac{e^{-2\beta}}{w_0^2}, \\ \langle(\Delta w)^2\rangle &= e^{-2\beta} w_0^2 \,. \end{aligned} \tag{A5}$$

Here β is a squeezing parameter, w_0 is

$$w_0 = \frac{1}{\sqrt{2\sigma_{\text{coh}}\sigma_{\text{cor}}}}\,, \tag{A6}$$

with σ_{coh} being the pump coherence time and σ_{cor} the biphoton correlation time.

We have assumed that the biphoton state has a Gaussian wave function. Reference [20] has proven that the variance of the frequency difference can be upper bounded from the Franson visibility via

$$\langle(\Delta w)^2\rangle \leq \frac{2(1-V_0)}{\Delta T^2}\,. \tag{A7}$$

Note that this bound applies only to high Franson visibilities, i.e., satisfying the assumptions made in [20]. By combining Eqs. (A5)–(A7), we arrive at the following upper bound of $d_w^{L_1}$

$$d_w^{L_1} = \sqrt{\frac{2}{\pi}}\frac{|\,\langle(\Delta t)\rangle\,|}{\delta} \leq \frac{\sigma_{\text{coh}}\sigma_{\text{cor}}}{\delta\Delta T}\sqrt{\frac{16(1-V_0)}{\pi}}\,, \tag{A8}$$

where δ is the time-bin duration selected in the protocol.

In the experiment, both the time-bin duration δ and the frame size (dimensionality) N_d were chosen in the data post-processing step by parsing the raw timing records into the desired symbol length. N_d was optimized to produce the maximal bits per photon: a larger N_d can produce more raw bits per sample, i.e., $-\log_2 c$ in Eq. A1, but it also increases $H_{\max}(W|\mathbf{B})$; hence, N_d was optimized to maximize $H_{\min}(T|\mathbf{E})$.

Appendix B: Quantification of accidental counts

In the RNG round, the detections made by system **A** can be due to either SPDC signal photons or be accidental counts. Given our SNSPDs' low dark-count rates, accidental counts are mainly due to the fluorescence (unpaired) photons in the SPDC's output. Here we quantify the SPDC output's unpaired-to-paired ratio.

The total number of photons/s N_{S} generated in the signal field by SPDC can be written as a sum of paired photons/s N_{SPDC} and fluorescence photons/s N_{F}:

$$N_{\text{S}} = N_{\text{SPDC}} + N_{\text{F}}. \tag{B1}$$

Given the overall detection efficiencies for the signal (η_{S}) and idler (η_{I}) photons, the singles rate (C_{S}) and coincidence rate (C_{SI}) can be written as

$$\begin{aligned} C_{\text{SI}} &= N_{\text{SPDC}}\eta_{\text{S}}\eta_{\text{I}}, \\ C_{\text{S}} &= N_{\text{S}}\eta_{\text{S}}. \end{aligned} \tag{B2}$$

In characterizing our entanglement source, a typical set of measurements yields the following values for the singles rate, coincidence rate, and efficiencies shown in Table II, and we obtain the unpaired-to-paired ratio for the signal field of the SPDC output

$$\frac{N_{\text{F}}}{N_{\text{SPDC}}} = 1.8\%. \tag{B3}$$

This result is consistent with reported ratios of 2% in previous measurements of PPKTP waveguide and PPKTP bulk crystal at the telecom wavelengths [21].

C_{SI}	C_{S}	η_{S}	η_{I}
420 kcoinc/s	850 kcounts/s	49.4%	50.3%

TABLE II: Experimental values of singles rate, coincidence rate, and efficiencies.

Statistical test	P-value	Result
Birthday Spacings [KS]	0.680563	success
Overlapping permutations	0.308246	success
Ranks of 31x31 matrices	0.450693	success
Ranks of 31x32 matrices	0.591037	success
Ranks of 6x8 matrices [KS]	0.448596	success
Bit stream test	0.06551	success
Monkey test OPSO	0.015300	success
Monkey test OQSO	0.098700	success
Monkey test DNA	0.098000	success
Count 1's in stream of bytes	0.461867	success
Count 1's in specific bytes	0.031698	success
Parking lot test [KS]	0.809513	success
Minimum distance test [KS]	0.915470	success
Random spheres test [KS]	0.902702	success
Squeeze test	0.350940	success
Overlapping sums test [KS]	0.795741	success
Runs test (up) [KS]	0.569616	success
Runs test (down) [KS]	0.248829	success
Craps test No. of wins	0.259975	success
Craps test throws/game	0.643893	success

TABLE III: DIEHARD. Data size is about 104 Mbits. 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. The test is successful if all final P-values satisfy $0.01 \leq P \leq 0.99$.

[1] X. Ma, X. Yuan, Z. Cao, B. Qi, and Z. Zhang, "Quantum random number generation," npj Quantum Information **2,** 16021 (2016).

[2] 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).

[3] J. F. Dynes, Z. L. Yuan, A. W. Sharpe, and A. J. Shields, "A high speed, postprocessing free, quantum random number generator," Appl. Phys. Lett. **93**, 031109 (2008)

[4] 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).

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

[6] C. Abellán, W. Amaya, M. Jofre, M. Curty, A. Acín, J. Capmany, V. Pruneri, and M. W. Mitchell, "Ultra-fast quantum randomness generation by accelerated phase diffusion in a pulsed laser diode," Opt. Express **22**, 1645 (2014).

[7] Z. L. Yuan, M. Lucamarini, J. F. Dynes, B. Frohlich, A. Plews, A. J. Shields, "Robust random number generation using steady-state emission of gain-switched laser diodes," Appl. Phys. Lett. **104**, 261112 (2014).

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

[9] D. Frauchiger, R. Renner, and M. Troyer, "True randomness from realistic quantum devices," arXiv:1311.4547 (2013).

[10] J. Y. Haw, S. M. Assad, A. M. Lance, N. H. Y. Ng, V. Sharma, P. K. Lam, and T. Symul, "Maximization of Extractable Randomness in a Quantum Random-Number Generator," Phys. Rev. Appl. **3**, 054004 (2015).

[11] R. Colbeck, "Quantum And Relativistic Protocols For Secure Multi-Party Computation," PhD thesis, University of Cambridge, (2006).

[12] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem," Nature **464,** 1021 (2010).

[13] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin, and P. G. Kwiat, "Detection-Loophole-Free Test of Quantum Nonlocality, and Applications," Phys. Rev. Lett. **111,** 130406 (2013).

[14] T. Lunghi, J. B. Brask, C. C. W. 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] T. Zhong, H. Zhou, R. D. Horansky, C. Lee, V. B. Verma, A. E. Lita, A. Restelli, J. C. Bienfang, R. P. Mirin, T. Gerrits, S. W. Nam, F. Marsili, M. D. Shaw, Z. Zhang, L. Wang, D. Englund, G. W. Wornell, J. H. Shapiro, and F. N. C. Wong, "Photon-efficient quantum key distribution using timeenergy entanglement with high-dimensional encoding," New J. Phys. **17,** 022002 (2015).

[16] J. D. Franson, "Bell inequality for position and time," Phys. Rev. Lett. **62,** 2205 (1989).

[17] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B. Scholz, M. Tomamichel, and R. F. Werner, "Continuous Variable Quantum Key Distribution: Finite-Key Analysis of Composable Security against Coherent Attacks," Phys. Rev. Lett. **109,** 100502 (2012).

[18] G. Vallone, D. G. Marangon, M. Tomasin, and P. Villoresi, "Quantum randomness certified by the uncertainty principle," Phys. Rev. A **90,** 052327 (2014).

[19] T. Zhong, and F. N. C. Wong, "Nonlocal cancellation of dispersion in Franson interferometry," Phys. Rev. A **88,** 020103(R) (2013).

[20] Z. Zhang, J. Mower, D. Englund, F. N. C. Wong, and J. H. Shapiro, "Unconditional Security of Time-Energy Entanglement Quantum Key Distribution Using Dual-Basis Interferometry," Phys. Rev. Lett. **112,** 120506 (2014).

[21] T. Zhong, F. N. C. Wong, A. Restelli, and J. C. Bienfang, "Efficient single-spatial-mode periodically-poled $KTiOPO_4$ waveguide source for high-dimensional entanglement-based quantum key distribution," Opt. Express **20,** 26868 (2012).

[22] Photon Spot: http://www..com/detectors.

[23] IDQ: http://www.idquantique.com; Picoquant: https://www.picoquant.com.

[24] Z. Cao, H. Zhou, X. Yuan, and X. Ma, "Source-independent quantum random number generation," Phys. Rev. X **6**, 011020 (2016).

[25] D. G. Marangon, G. Vallone, and P. Villoresi, "Source-device-independent Ultra-fast Quantum Random Number Generation," arXiv:1509.07390 (2015).

[26] T. Ikuta and H. Takesue, "Enhanced violation of the Collins-Gisin-Linden-Massar-Popescu inequality with optimized time-bin-entangled ququarts," Phys. Rev. A **93**, 022307 (2016).

[27] A. Cabello, A. Rossi, G. Vallone, F. De Martini and P. Mataloni, "Proposed Bell Experiment with Genuine Energy-Time Entanglement," Phys. Rev. Lett. **102**, 040401 (2009).

[28] M. Tomamichel and R. Renner, "Uncertainty relation for smooth entropies," Phys. Rev. Lett. **106**, 110506 (2011).