Self-Testing Quantum Random Number Generator

Tommaso Lunghi,[1] Jonatan Bohr Brask,[2] Charles Ci Wen Lim,[1] Quentin Lavigne,[1] Joseph Bowles,[2]
Anthony Martin,[1] Hugo Zbinden,[1] and Nicolas Brunner[2]
[1]*Group of Applied Physics, Université de Genève, 1211 Genève, Switzerland*
[2]*Département de Physique Théorique, Université de Genève, 1211 Genève, Switzerland*

(Received 14 November 2014; published 15 April 2015)

The generation of random numbers is a task of paramount importance in modern science. A central problem for both classical and quantum randomness generation is to estimate the entropy of the data generated by a given device. Here we present a protocol for self-testing quantum random number generation, in which the user can monitor the entropy in real time. Based on a few general assumptions, our protocol guarantees continuous generation of high quality randomness, without the need for a detailed characterization of the devices. Using a fully optical setup, we implement our protocol and illustrate its self-testing capacity. Our work thus provides a practical approach to quantum randomness generation in a scenario of trusted but error-prone devices.

DOI: 10.1103/PhysRevLett.114.150501 PACS numbers: 03.67.Ac, 42.50.Ex

Given the importance of randomness in modern science and beyond, e.g., for simulation algorithms and for cryptography, an intense research effort has been devoted to the problem of extracting randomness from quantum systems. Devices for quantum random number generation (QRNG) are now commercially available. All of these schemes work essentially according to the same principle, exploiting the randomness of quantum measurements. A simple realization consists in sending a single photon on a 50/50 beam splitter and detecting the output path [1–3]. Other designs were developed, based on measuring the arrival time of single photons [4–7], the phase noise of a laser [8–10], vacuum fluctuations [11,12], and even mobile phone cameras [13].

A central issue in randomness generation is the problem of estimating the entropy of the bits that are generated by a device, i.e., how random is the raw output data. When a good estimate is available, appropriate postprocessing can be applied to extract true random bits from the raw data (via a classical procedure termed randomness extractor [14]). However, poor entropy estimation is one of the main weaknesses of classical RNG [15], and can have important consequences. In the context of QRNG, entropy estimates for specific setups were recently provided using sophisticated theoretical models [16,17]. Nevertheless, this approach has several drawbacks. First, these techniques are relatively cumbersome, requiring estimates for numerous experimental parameters which may be difficult to precisely assess in practice. Second, each study applies to a specific experimental setup, and cannot be used for other implementations. Finally, it offers no real-time monitoring of the quality of the RNG process, hence no protection against unnoticed misalignment (or even failures) of the experimental setup.

It is therefore highly desirable to design QRNG techniques which can provide a real-time estimate of the output entropy. An elegant solution is provided by the concept of device-independent QRNG [18,19], where randomness can be certified and quantified without relying on a detailed knowledge of the functioning of the devices used in the protocol. Nevertheless, the practical implementation of such protocols is extremely challenging as it requires the genuine violation of Bell's inequality [19,20]. Alternative approaches were proposed [21] but their experimental implementation suffers from loopholes [22]. More recently, an approach based on the uncertainty principle was proposed but requires a fully characterized measurement device [23].

Here, we present a simple and practical protocol for self-testing QRNG. Based on a prepare-and-measure setup, our protocol provides a continuous estimate of the output entropy. Our approach requires only a few general assumptions about the devices (such as quantum systems of bounded dimension) without relying on a detailed model of their functioning. This setting is relevant to real-world implementations of randomness generation, and is well adapted to a scenario of trusted but error-prone providers, i.e., a setting where the devices used in the protocol are not actively designed to fool the user, but where implementation may be imperfect. The key idea behind our protocol is to certify randomness from a pair of incompatible quantum measurements. As the incompatibility of the measurements can be directly quantified from experimental data, our protocol is self-testing. That is, the amount genuine quantum randomness can be quantified directly from the data, and can be separated from other sources of randomness such as fluctuations due to technical imperfections. We implemented this scheme with standard technology, using a single photon source and fibered telecommunication components. We implement the complete QRNG protocol, achieving a rate 23 certified random bits per second, with 99% confidence.

Protocol.—Our protocol, sketched in Fig. 1, uses two devices which, respectively, prepare and measure an

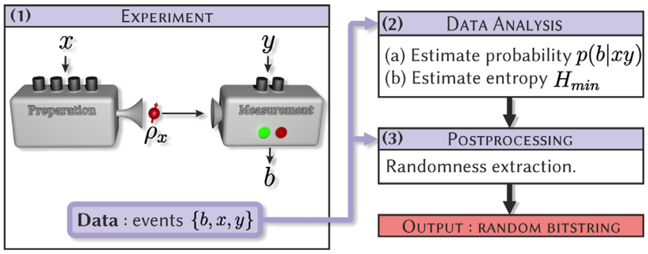
(1) Experiment
x
y
Preparation
Measurement
ρ_x
b
Data : events $\{b,x,y\}$
(2) Data Analysis
(a) Estimate probability $p(b|xy)$
(b) Estimate entropy H_{min}
(3) Postprocessing
Randomness extraction.
Output : random bitstring

FIG. 1 (color online). Sketch of the protocol. The self-testing QRNG protocol consists of three distinct steps. (1) First, an experiment is performed where, in each round, the user chooses a preparation x and a measurement y, and obtains an outcome b. (2) From the raw data, the distribution $p(b|x,y)$ can be estimated leading to an estimate for the value of the witness W, from which the entropy of the raw data can be quantified. (3) Based on the entropy bound, appropriate postprocessing of the raw data is performed, in order to extract the final random bit string.

uncharacterized qubit system. In each round of the protocol, the observer chooses settings among four possible preparations, $x = 0, 1, 2, 3$, and two measurements $y = 0, 1$, resulting in a binary outcome $b = \pm 1$. To model imperfections, we represent the internal state of each device by a random variable—λ for the preparation device and μ for the measurement device—which are unknown to the observer. As we work in a scenario where the devices are not maliciously conspiring against the user, we assume the devices to be independent, i.e., $p(\lambda,\mu) = q(\lambda)r(\mu)$, where $\int d\lambda q(\lambda) = \int d\mu r(\mu) = 1$.

In each round of the experiment, the preparation device emits a qubit state ρ_x^λ which depends on the setting x and on the internal state λ. Similarly, the measurement device performs a measurement M_y^μ. Thus the distributions of λ and μ determine the distributions of the prepared states and the measurements. As the observer has no access to the variables λ and μ, he will observe

$$p(b|x,y) = \int d\lambda q(\lambda) \int d\mu r(\mu) p(b|x,y,\lambda,\mu)$$
$$= \mathrm{Tr}\left(\rho_x \frac{\mathbb{1} + bM_y}{2}\right) = \frac{1}{2}(1 + b\vec{S}_x \cdot \vec{T}_y), \quad (1)$$

where

$$\rho_x = \int d\lambda q(\lambda)\rho_x^\lambda = \frac{1}{2}(\mathbb{1} + \vec{S}_x \cdot \vec{\sigma}), \quad (2)$$

$$M_y = \int d\mu r(\mu) M_y^\mu = \vec{T}_y \cdot \vec{\sigma}. \quad (3)$$

Here, $\vec{S}_x$ and $\vec{T}_y$ denote the Bloch vectors of the (average) states and measurements, and $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector of Pauli matrices.

The task of the observer is to estimate the amount of genuine quantum randomness generated in this setup, based only on the observed distribution $p(b|x,y)$. This is a nontrivial task as the apparent randomness of the distribution $[0 < p(b|x,y) < 1]$ can have different origins. On the one hand, it could be genuine quantum randomness. That is, if in a given round of the experiment, the state ρ_x^λ is not an eigenstate of the measurement operator M_y^μ, then the outcome b cannot be predicted with certainty, even if the internal states λ and μ are known, i.e., $0 < p(b|x,y,\lambda,\mu) < 1$. On the other hand, the apparent randomness may be due to technical imperfections, that is, to fluctuations of the internal states λ and μ. Consider the following example: The preparation device emits the states $\rho_x^{\lambda=0} = |0\rangle\langle 0|$ and $\rho_x^{\lambda=1} = |1\rangle\langle 1|$ with $q(\lambda = 0, 1) = 1/2$. For a measurement of the observable $M_y = \hat{z} \cdot \vec{\sigma}$, one obtains that $p(b|x,y) = 1/2$. However, these data clearly contain no quantum randomness, since the outcome b can be perfectly guessed if the internal state λ is known.

Our protocol allows the observer to separate quantum randomness from the randomness due to technical noise. The key technical tool of our protocol is a function recently presented in [24], which works as a "dimension witness." Given data $p(b|x,y)$, the quantity

$$W = \begin{vmatrix} p(1|0,0) - p(1|1,0) & p(1|2,0) - p(1|3,0) \\ p(1|0,1) - p(1|1,1) & p(1|2,1) - p(1|3,1) \end{vmatrix} \quad (4)$$

captures the quantumness of the preparation and measurements. Specifically, if the preparations are classical (i.e., there exists a basis in which all states ρ_x^λ are diagonal), one has that $W = 0$, while a generic qubit strategy achieves $0 \leq W \leq 1$ [24]. $W > 0$ guarantees that the measurements performed by Bob are incompatible (see [25]) and since it is then impossible to simultaneously assign deterministic outcomes to them, this enables us to bound the guessing probability and certify randomness. Given x, y, and knowledge of the internal states λ, μ, the best guess for b is given by $\max_b p(b|x,y,\lambda,\mu)$. Assuming uniformly distributed x and y, the average probability of guessing b fulfils the following inequality (see [25]):

$$p_{\text{guess}} = \frac{1}{8} \sum_{x,y,\lambda,\mu} q_\lambda r_\mu \max_b p(b|x,y,\lambda,\mu)$$
$$\leq \frac{1}{2}\left(1 + \sqrt{\frac{1 + \sqrt{1 - W^2}}{2}}\right). \quad (5)$$

Therefore, the guessing probability can be upper bounded by a function of W, which can be determined directly from the data $p(b|x,y)$. Finally, to extract random bits from the raw data, we use a randomness extraction procedure. The number of random bits that can be extracted per experimental run is given by the min-entropy $H_{\min} = -\log_2 p_{\text{guess}}$ [27]. Hence $H_{\min}$ is the relevant parameter for determining how the raw data must be postprocessed. Note that randomness can be extracted for any $W > 0$, since $p_{\text{guess}} < 1$ in this case.

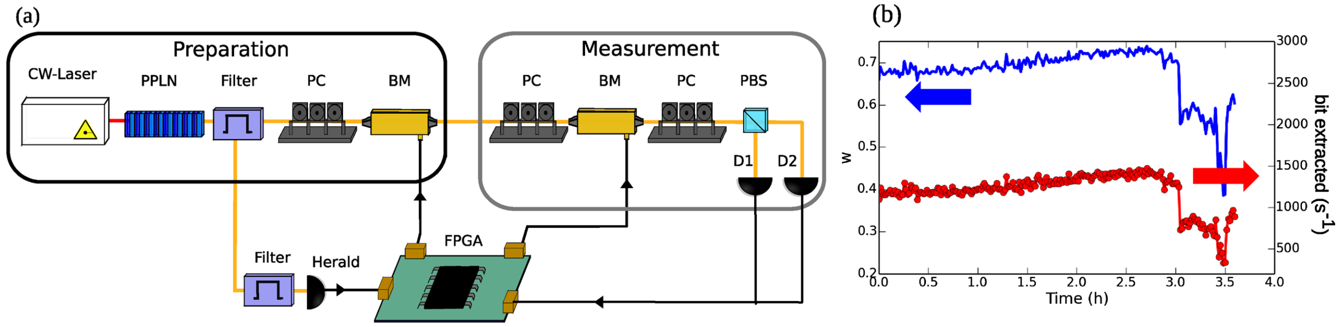

FIG. 2 (color online). Implementing the self-testing QRNG. (a) Experimental setup. (b) Real-time evolution of the witness value W (blue) and randomness generation rate (bits extracted per second; red). After 3 h, the air conditioning in the laboratory is switched off, which leads to misalignment of the optical components. In turn, this leads to a significant drop of the witness value W and corresponding entropy.

The maximal value of $W = 1$ can be reached using the set of preparations and measurements: $\vec{S}_0 = -\vec{S}_1 = \vec{T}_0 = \hat{z}$ and $\vec{S}_2 = -\vec{S}_3 = \vec{T}_1 = \hat{x}$, which correspond to the BB84 QKD protocol [28]. In this case, we can certify randomness with min-entropy $H_{\min} \simeq 0.2284$. Using other preparations and measurements, e.g., if the system is noisy or becomes misaligned, one will typically obtain $0 < W < 1$. Nevertheless, for any value $W > 0$, randomness can be certified, and the corresponding min-entropy can be estimated using Eq. (5). Our protocol is therefore self-testing, since the evaluation of W allows quantifying the amount of randomness in the data. In turn, this allows one to perform adapted postprocessing in order to finally extract random bits.

To conclude this section, we discuss the assumptions which are required in our protocol: (i) *Choice and distribution of settings.*—The devices make no use of any prior information about the choice of settings x and y. (ii) *Internal states of the devices are independent and identically distributed (i.i.d).*—The distributions $q(\lambda)$ and $r(\mu)$ do not vary between experimental rounds. (iii) *Independent devices.*—The preparation and measurement devices are independent, in the sense that $p(\lambda, \mu) = q(\lambda)r(\mu)$. (iv) *Qubit channel capacity.*—The information about the choice of preparation x retrieved by the measurement device (via a measurement on the mediating particle) is contained in a two-dimensional quantum subspace (a qubit).

Assumptions (i) and (iii) are arguably rather natural in a setting where the devices are produced without malicious intent. They concern the independence of devices used in the protocol, namely the preparation and measurement devices, and the choice of settings. When these are produced by trusted (or simply different) providers, it is reasonable to assume that there are no (built-in) preestablished correlations between the devices and that the settings x, y can be generated independently, e.g., using a pseudo RNG. Assumptions (ii) and (iv) are stronger, and will have to be justified for the particular implementation at hand. The content of assumption (ii) is essentially that the devices are memoryless (internal states do not depend on previous events). We believe this assumption can likely be weakened, since randomness can in fact be guaranteed in the presence of certain memory effects, in particular, the experimentally relevant afterpulsing effect (see [25]). Finally, note that assumption (iv) restricts the amount of information about x that is retrieved by the measuring device (via a measurement on the mediating particle), but not the information about x contained in the mediating particle itself. In other words, it might be the case that information about x leaks out from the preparation device via side channels, but we assume that these side channels are not maliciously exploited by the measurement device.

Experiment.—We implemented the above protocol using a fully guided optical setup [see Fig. 2(a)]. The qubit preparations are encoded in the polarization state of single photons, generated via a heralded single-photon source based on a continuous wave (CW) spontaneous parametric down conversion (SPDC) process in a periodically poled lithium niobate (PPLN) waveguide [29]. The idler photon is detected with a ID220 free-running InGaAs/InP single-photon detector (SPD) (herald) with 20% detection efficiency and 20 μs dead time. The polarization is rotated using a polarization controller (PC) and an electro-optical birefringence modulator (BM) based on a lithium niobate waveguide phase modulator. The preparations $x = \{0, 1, 2, 3\}$ correspond, respectively, to the diagonal (D), antidiagonal (A), circular right (R), and circular left (L) polarization states. For the measurement device, polarization measurements are done using a BM and a PC followed by a polarization beam splitter and two ID210 InGaAs/InP SPDs (with a 1.5 ns gate and 25% detection efficiency) triggered by a detection at the heralding detector. The measurements $y = \{0, 1\}$ correspond, respectively, to the $\{D, A\}$ basis and the $\{R, L\}$ basis. The number of photon pairs generated by the SPDC source is set to obtain a count rate at the heralding detector of about 30 kHz, which corresponds to a probability of single photon emission of $p_1 = 6.5 \times 10^{-4}$ per gate, and a two photon emission

$p_2 = p_1^2/2 = 2.1 \times 10^{-7}$ per gate. A field-programmable-gate-array board (FPGA) continuously generates sequences of three pseudorandom bits. Upon successful heralding, these three bits are used to choose (x, y). Finally, the FPGA records the outcome b (whether each ID210 detector has clicked or not).

We briefly discuss to which extent the assumptions of the protocol fit to our implementation. First, the choice of preparation and measurement, x and y, are made by the FPGA using a linear-feedback shift register pseudo RNG [30]. This RNG provides a deterministic cyclic function sampled by the heralding detector. Since the sampling is asynchronous with respect to the RNG rate, the output is uniform and (i) is fulfilled. The BMs are separated spatially by 1 m, their temperature is controlled independently, and the voltages are applied with independent electronic circuits. Any cross talk between them, e.g., due to stray electric fields, can be safely neglected; hence, (iii) is also satisfied. Concerning assumption (ii), we evaluate the distribution $p(b|x, y)$ after every minute of acquisition. Therefore, we need to consider memory effects with time characteristics shorter than one minute. Two main effects should be considered: charge accumulation in the birefringence modulator, and afterpulsing in the detectors, which is a common issue in standard QRNG approaches [4,16]. Importantly, our protocol is robust to afterpulsing (see [25]). Charge effects in the modulator are relevant only for modulation slower than 1 Hz [31]. Finally, the qubit assumption (iv) is arguably the most delicate one. As the choice of preparation x is encoded in the polarization of a single photon, (iv) seems justified. However, a small fraction of heralded events corresponds to multiphoton pulses, in which (iv) is not valid. To take these events into account, we extend our theoretical analysis (see [25]). We show that quantum randomness can still be guaranteed even when (iv) is not fulfilled in all experimental events, provided that the fraction of events violating (iv) can be bounded and is small enough compared to the total number of successful events. To verify this assumption, the probability of single and multiphoton pulses must be properly calibrated. For our single-photon source, the ratio of multiphoton events vs heralds is given by $\sim p_1/2 = 3.25 \times 10^{-4}$, and our method can be applied.

We ran the experiment estimating W for the data accumulated each minute. As discussed in [25], the estimation of W considers finite-size effects and the size of the randomness extractor is determined based on the value of W [16,32]. In the best conditions, our setup generates about 402 bits/s of raw data (before the extractor). The witness corresponds to a value of $W = 0.76$. After extraction, we get final random bits at a rate of 23 bits/s with a confidence of 99%. Note that the confidence level is set when accounting for finite size effects; a higher confidence can be chosen at the expense of a lower rate. Note also that this rate is limited by the slow repetition rate of the experiment (limited by the dead time of the heralding detector) and by the losses in the optical implementation (channel transmission is ~8%; total efficiency ~2%). Figure 2(b) shows the estimated value of W over 3.5 h and the rate at which the final random bits are generated. To demonstrate the self-testing capacity of our protocol, we switched off the air conditioning in the room after 3 h. This impacts the alignment of the setup. As can be seen from Fig. 2(b), the witness value W drops, reflecting the fact that the distributions of internal states [$q(\lambda)$ and $r(\mu)$] changed. In turn, this forces us to perform more postprocessing, resulting in a lower randomness generation rate. Nevertheless, the quality of the final random bits is still guaranteed. This shows that our setup can warrant the generation of high quality randomness, without active stabilization or precise modeling of the impact of the temperature increase.

The quality of the generated randomness can be assessed by checking for patterns and correlations in the extracted bits. We performed standard statistical tests, as defined by NIST, and although not all tests could be performed due to the small size of the sample, all performed tests were successful (see [25]). We stress that these tests do not constitute a proof of randomness (which is impossible); however, failure to pass any of them would indicate the presence of correlations among the output bits.

Finally, we comment on the influence of losses. In the above analysis, we discarded inconclusive events in which the photon was not detected at the measuring device, although the emission of a single photon was heralded by the source. Therefore, our analysis is subject to an additional assumption, namely, that of fair sampling, which we believe is rather natural in the case of nonmalicious devices. Note, however, that this is not necessary strictly speaking, as our protocol is in principle robust to arbitrarily low detection efficiency [24]. Performing the data analysis without the fair-sampling assumption (in which case the inconclusive events are attributed the outcome -1) we obtain witness values of $W \sim 1.5 \times 10^{-4}$, corresponding to $H_{\min} \sim 2.0 \times 10^{-9}$. In this case, the rate for generating random bits drops considerably to 6×10^{-5} bits/s, but importantly does not vanish. Hence, our setup can be used to certify randomness without requiring the fair-sampling assumption. We note that even a small increase in efficiency would lead to a large improvement in rate. E.g., an increase from our current 2% to 10% would already give ~0.04 bits/s while an overall efficiency of 50% would be enough to reach 23 bits/s without postselection, equal to our current postselected rate.

Conclusion.—We have presented a protocol for self-testing QRNG, which allows for real-time monitoring of the entropy of the raw data. This allows adapting the randomness extraction procedure in order to continuously generate high quality random bits. Using a fully optical guided implementation, we have demonstrated that our protocol is practical and efficient, and illustrated its self-testing capacity. Our work thus provides an approach to QRNG, which can be viewed as intermediate between the standard (device-dependent) approach and the device-independent one.

Compared to the device-dependent approach, our protocol delivers a stronger form of security requiring less characterization of the physical implementation, at the price of a reduced rate compared to commercial QRNGs such as ID Quantique QUANTIS which reaches 4 Mbits/s. A fully device-independent approach [18,19], on the other hand, offers even stronger security [in particular assumptions (ii)–(iv) can be relaxed, hence offering robustness to side channels and memory effects], but its practical implementation is extremely challenging. Proof-of-principle experiments require state-of-the-art setups but could achieve only very low rates [19,20]. Our approach arguably offers a weaker form of security, but can be implemented with standard technology. Our work considers a scenario of trusted but error-prone devices, which we believe to be relevant in practice.

We thank Antonio Acin, Stefano Pironio, Valerio Scarani, and Eric Woodhead for discussions; Raphael Houlmann and Claudio Barreiro for technical support; Batelle and ID Quantique for providing the PPLN waveguide. We acknowledge financial support from the Swiss National Science Foundation (Grant No. PP00P2_138917, Starting Grant DIAQ, and QSIT), SEFRI (COST action MP1006), and the EU project SIQS.

T. L. and J. B. B. contributed equally to this work.

Note added.—After submission of this work, several related works have appeared [33–35].

[1] J. Rarity, P. Owens, and P. Tapster, J. Mod. Opt. **41**, 2435 (1994).

[2] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, J. Mod. Opt. **47**, 595 (2000).

[3] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, Rev. Sci. Instrum. **71**, 1675 (2000).

[4] J. F. Dynes, Z. L. Yuan, A. W. Sharpe, and A. J. Shields, Appl. Phys. Lett. **93**, 031109 (2008).

[5] M. Wahl, M. Leifgen, M. Berlin, T. Rhlicke, H.-J. Rahn, and O. Benson, Appl. Phys. Lett. **98**, 171105 (2011).

[6] Y.-Q. Nie, H.-F. Zhang, Z. Zhang, J. Wang, X. Ma, J. Zhang, and J.-W. Pan, Appl. Phys. Lett. **104**, 051110 (2014).

[7] M. Stipčević and B. M. Rogina, Rev. Sci. Instrum. **78**, 045104 (2007).

[8] B. Qi, Y.-M. Chi, H.-K. Lo, and L. Qian, Opt. Lett. **35**, 312 (2010).

[9] A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, OowadaIsao, T. Kurashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis, Nat. Photonics **2**, 728 (2008).

[10] C. Abellán, W. Amaya, M. Jofre, M. Curty, A. Acín, J. Capmany, V. Pruneri, and M. W. Mitchell, Opt. Express **22**, 1645 (2014).

[11] C. Gabriel, C. Wittmann, D. Sych, R. Dong, W. Mauerer, U. L. Andersen, C. Marquardt, and G. Leuchs, Nat. Photonics **4**, 711 (2010).

[12] T. Symul, S. M. Assad, and P. K. Lam, Appl. Phys. Lett. **98**, 231103 (2011).

[13] B. Sanguinetti, A. Martin, H. Zbinden, and N. Gisin, Phys. Rev. X **4**, 031056 (2014).

[14] N. Nisan and A. Ta-Shma, J. Comput. Syst. Sci. **58**, 148 (1999).

[15] Y. Dodis, D. Pointcheval, S. Ruhault, D. Vergniaud, and D. Wichs, in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security CCS '13* (ACM, New York, 2013), p. 647.

[16] D. Frauchiger, R. Renner, and M. Troyer, arXiv:1311.4547.

[17] X. Ma, F. Xu, H. Xu, X. Tan, B. Qi, and H.-K. Lo, Phys. Rev. A **87**, 062327 (2013).

[18] R. Colbeck, Ph.D. thesis, Trinity College, University of Cambridge [arXiv:0911.3814].

[19] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature (London) **464**, 1021 (2010).

[20] 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, Phys. Rev. Lett. **111**, 130406 (2013).

[21] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **84**, 034301 (2011); H.-W. Li, M. Pawłowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **85**, 052308 (2012).

[22] M. Dall'Arno, E. Passaro, R. Gallego, M. Pawlowski, and A. Acin, Quantum Inf. Comput. **15**, 0037 (2015).

[23] G. Vallone, D. G. Marangon, M. Tomasin, and P. Villoresi, Phys. Rev. A **90**, 052327 (2014).

[24] J. Bowles, M. T. Quintino, and N. Brunner, Phys. Rev. Lett. **112**, 140407 (2014).

[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.150501, which includes Ref. [26], for details of the proof of randomness and discussion of assumptions, afterpulsing, multiphoton events, and statistical tests.

[26] W. Hoeffding, J. Am. Stat. Assoc. **58**, 13 (1963).

[27] R. Koenig, R. Renner, and C. Schaffner, IEEE Trans. Inf. Theory **55**, 4337 (2009).

[28] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 27.

[29] S. Tanzilli, A. Martin, F. Kaiser, M. De Micheli, O. Alibart, and D. Ostrowsky, Laser Photonics Rev. **6**, 115 (2012).

[30] P. Alfke, *Efficient Shift Registers, LFSR Counters, and Long Pseudo-Random Sequence Generators* (Xilinx Inc., San Jose, CA, 1996).

[31] E. Wooten, K. Kissa, A. Yi-Yan, E. Murphy, D. Lafaw, P. Hallemeier, D. Maack, D. Attanasio, D. Fritz, G. McBrien, and D. Bossi, IEEE J. Sel. Top. Quantum Electron. **6**, 69 (2000).

[32] M. Troyer and R. Renner, A Randomness Extractor for the Quantis Device, http://www.idquantique.com/images/stories/PDF/quantis-random-generator/quantis-rndextract-techpaper.pdf.

[33] M. W. Mitchell, C. Abellan, and W. Amaya, Phys. Rev. A **91**, 012314 (2015).

[34] G. Cañas *et al.*, arXiv:1410.3443.

[35] J. Y. Haw *et al.*, arXiv:1411.4512.