A 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 Brunner2

1 Group of Applied Physics, University of Geneva, 1211 Genève, Switzerland
2 Département de Physique Théorique, Université de Genève, 1211 Genève, Switzerland

A central issue in randomness generation is to estimate the entropy of the output data generated by a given device. Here we present a protocol for self-testing quantum random number generation, in which the entropy of the raw data can be monitored in real-time. In turn, this allows the user to adapt the randomness extraction procedure, in order to continuously generate high quality random bits. Using a fully optical implementation, we demonstrate that our protocol is practical and efficient, and illustrate its self-testing capacity.

I. INTRODUCTION

The generation and certification of random numbers is a task of paramount importance in modern society, e.g. for running simulation algorithms as well as for cryptography. In this context, quantum mechanical systems are highly relevant, given that intrinsic randomness is a distinctive feature of quantum theory. Over the last two decades, an intense research effort has been devoted to the problem of extracting randomness from quantum systems, and devices for quantum random number generation (QRNG) are now commercially available. All these schemes work essentially according to the same principle, exploiting the randomness of a particular quantum measurement. A simple realization consists in sending a single photon on a 50/50 beam-splitter and detecting the output path [1, 2]. Other designs were also developed, based on measuring the arrival time of single photons [3–6], the phase noise of a laser [7–9], vacuum fluctuations [10, 11], and even mobile phone cameras [12].

A central issue in randomness generation (classical and quantum) 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 post-processing can be applied to extract true random bits from the raw data (via a classical procedure termed randomness extractor [13]). However, poor entropy estimation is one of the main weaknesses of classical RNG [14], and can have important consequences. In the context of QRNG, entropy estimates for specific setups were recently provided using sophisticated theoretical models [15, 16]. 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 QNRG techniques which can provide a real-time estimate of the output entropy. An elegant solution is provided by the concept of device-independent QRNG [17, 18], 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 [18, 19]. Alternative approaches were proposed [20] but their experimental implementation suffers from loopholes [21]. More recently, an approach based on the uncertainty principle was proposed but requires a fully characterized measurement device [22].

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 independence of the preparation and measurement devices, and 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. This protocol can be implemented with standard technology, as we demonstrate using a single photon source and fibered telecommunications components. We implement the complete QRNG protocol, achieving a rate 23 certified random bits per second, with 99% confidence.

* These authors contributed equally to this work.
As we work in a scenario where the devices are not maliciously conspiring against the user, we will 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 run of the experiment, the preparation device emits a qubit state $\rho_x^\lambda$ which depends on the setting $x$ and on the internal state $\lambda$. Similarly, the measurement device performs a measurement $M_y^\mu$. As the observer has no access to the variables $\lambda$ and $\mu$, he will observe the distribution
\[
p(b|x,y) = \int d\lambda q(\lambda) \int d\mu r(\mu) p(b|x,y,\lambda,\mu) = \operatorname{Tr}(\rho_x^{\lambda} + bM_y^\mu) = \frac{1}{2} (1 + b\vec{S}_x \cdot \vec{T}_y) \quad (1)
\]
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 data $p(b|x,y) < 1$ can have different origins. On the one hand, it could be genuine quantum randomness. That is, if in a given run of the experiment, the state $\rho_x^\lambda$ is not an eigenstate of the measurement operator $M_y^\mu$, then the outcome $b$ cannot be predicted with certainty, even if the internal states $\lambda$ and $\mu$ 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 $\lambda$ and $\mu$. 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 $M_y = \vec{z} \cdot \vec{\sigma}$, one obtains that $p(b|x,y) = 1/2$. However, this data clearly contains no quantum randomness, since the outcome $b$ can be perfectly guessed if the internal state $\lambda$ is known.

Our protocol allows the observer to separate quantum randomness from the data due to technical noise. It enables the observer to certify a certain amount of quantum randomness without knowing the internal states $\lambda$ and $\mu$, and without detailed knowledge of the preparation and measurements used in the protocol. The key technical tool of our protocol is a function recently presented in [23], which works as a 'dimension witness'. Given data $p(b|x,y)$, the quantity
\[
W = \left| \begin{array}{ccc}
p(1|0,0) - p(1|1,0) & p(1|2,0) - p(1|3,0) \\
p(1|0,1) - p(1|1,1) & p(1|2,1) - p(1|3,1)
\end{array} \right| \quad (4)
\]
captures the quantumness of the preparation and measurements. Specifically, if the preparations are classical
(i.e. there exist a basis in which all states $\rho^b_\lambda$ are diagonal), one has that $W = 0$. On the contrary, a generic qubit strategy achieves $0 \leq W \leq 1$ [23]. Notably, the quantity $W$, which can be estimated directly from the data, characterizes the incompatibility of the two measurements performed by Bob. Specifically, one has that

$$\int d\mu \rho \| [M^0_b, M^1_b] \| \geq 2W$$

(5)

Since it is impossible to simultaneously assign deterministic outcomes to incompatible quantum measurements, this in turn enables us to bound the guessing probability. Given settings $x, y$, and knowledge of the internal states $\lambda, \mu$, the best guess for $b$ is given by

$$\max_b p(b|x, y, \lambda, \mu).$$

(6)

Assuming uniformly distributed inputs $x$ and $y$, the average probability of guessing $b$ can be shown to fulfill the following inequality (see App. A)

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

(7)

Therefore the guessing probability can be upper-bounded by a function of $W$, the value of which can be determined directly from the data $p(b|x, y)$. To extract random bits from the data, we use a randomness extraction procedure. The number of random bits that can be extracted per run of the experiment is given by the min-entropy $H_{\text{min}} = -\log_2 p_{\text{guess}}$ [23]. Hence $H_{\text{min}}$ is the relevant parameter for determining how the raw data must be post-processed in order to obtain the final random bits. Note that randomness can be extracted for any $W > 0$, since $p_{\text{guess}} < 1$ in this case.

The maximal value of $W = 1$ can be reached using the following set of preparations and measurements: $S_0 = -S_1 = T_0 = \hat{z}$ and $S_2 = -S_3 = T_1 = \hat{x}$, which correspond to the BB84 QKD protocol [23]. In this case, we can certify randomness with min-entropy $H_{\text{min}} \simeq 0.2284$. 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 equation (7). Our protocol is therefore self-testing, since the evaluation of the parameter $W$ allows to quantify the amount of quantum randomness contained in the data. In turn, this allows one to perform adapted post-processing in order to finally extract random bits.

To conclude this section, we enumerate and 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) **Independent devices.** The preparation and measurement devices have no initial shared randomness, i.e. $p(\lambda, \mu) = q(\lambda)r(\mu)$.

(iii) **Independent and identically distributed (i.i.d) events.** The internal states $\lambda$ and $\mu$ of the devices do not depend on previous events.

(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 2-dimensional quantum subspace (a qubit).

Assumptions (i) and (ii) are arguably rather natural in a setting where the devices are produced without malicious intent, but are subject to imperfections. They concern the independence of different 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) pre-established correlations between the devices, and in particular that the settings $x, y$ can be generated independently, e.g. using a pseudo-RNG.

Assumptions (iii) and (iv) are stronger, and will have to be justified for the particular implementation at hand. The content of assumption (iii) is essentially that the devices are memoryless. Note however that we believe this assumption can be weakened, since randomness can in fact be guaranteed in the presence of certain memory effects, in particular the experimentally relevant afterpulsing effect (see App. [23]). 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.

### III. EXPERIMENTAL RESULTS

We implemented the above protocol using a fully-guided optical setup (see Fig. 3). The qubit preparations are encoded in the polarization state of single photons, generated via a heralded single-photon source based on spontaneous parametric down-conversion process in a periodically poled lithium niobate waveguide [26]. Specifically, the preparations $x = \{0, 1, 2, 3\}$ correspond respectively to the diagonal (D), anti-diagonal (A), circular right (R) and circular left (L) polarization states. The measurements $y = \{0, 1\}$ correspond respectively to the $\{D, A\}$ basis and the $\{R, L\}$ basis.

Before showing the results we discuss to which extent the assumptions of the protocol fit to our implementation. First, the choice of preparation and measurement
FIG. 2. Self-testing QRNG: the preparations are encoded in the polarization of single photons, generated via a heralded single-photon source based on spontaneous parametric down-conversion process in a periodically poled lithium niobate waveguide (PPLN). The photon pairs are deterministically splitted and filtered using Dense-Wavelength Multiplex filters (Filter). The idler photon is detected with an ID220 free-running InGaAs/InP SPD (herald) with 20% of detection efficiency and 20 µs of dead time. Given the choice of preparation, \( x \), the polarization of the signal photon is then rotated using a polarization controller (PC) and an electro-optical birefringence modulator (BM) based on a lithium niobate waveguide phase modulator. The different states are prepared by applying a voltage to the BM. Similarly to the preparation stage, the measurement device is based on a BM and a PC followed by a polarization beam splitter (PBS). The desired measurement is performed by applying a voltage to the BM. The photons exiting from each port of the PBS are detected using two ID210 InGaAs/InP single-photon detectors (SPD) triggered by a detection at the heralding detector. The two ID210 detectors are gated with a 1.5 ns gate width with 25% of detection efficiency. Note that 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 emission of \( p_1 = 6.5 \times 10^{-4} \) per gate. This corresponds to a probability of having a two photon pair equal to \( p_2 = p_1^2/2 = 2.1 \times 10^{-7} \) per gate. A Field-Programmable-Gate-Array board (FPGA) generates the choices of measurement settings \( (x, y) \), by controlling the voltages applied to the BM. The FPGA continuously generates sequences of 3 pseudo-random bits. When the heralding detector fires, the 3 bits generated by the FPGA are used to set the BM’s voltages via two Digital-to-Analog Converters (DACs). Finally, the FPGA records the choice of preparation \( x \) and measurement \( y \), and the measurement outcome \( b \) (more precisely whether each ID210 detector has clicked or not).

basis, \( x \) and \( y \), are made by the FPGA using a linear-feedback shift register pseudo-RNG [27]. This RNG consists of 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. Then, the birefringence modulators are separated spatially by 1 m, their temperature is controlled independently, and the voltages are applied with independent electronic circuits. Any cross-talk between the birefringence modulators, e.g. due to stray electric fields, can be safely neglected, hence (ii) is also fulfilled. Concerning assumption (iii), we note that 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 1 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 [8][14]. Importantly, our protocol is robust to afterpulsing. As we show in App. [8], afterpulsing reduces the visibility and is accounted for in the randomness estimation. Charge effects in the modulator are relevant only for modulation slower than 1 Hz [28]. 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 multi-photon pulses, in which (iv) is not valid. To take these events into account, we must extend our theoretical analysis. Specifically, we develop a method for taking multi-photon pulses into account (see App. [8]). 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 multi-photon pulses must be properly calibrated. For our single-photon source, the ratio of multi-photon events vs heralds is given by \( \sim p_1/2 = 3.25 \times 10^{-4} \), and our method can be successfully applied.

We now present the results. We ran the experiment for 22 hours continuously, estimating the value of the quantity \( W \) for the data accumulated each minute. Notably, the estimation of \( W \) considers the finite-size effects as dis-
FIG. 3. **Experimental results.** (a) Real-time evolution of the witness value $W$ (blue) and randomness generation rate (bits extracted per second; red) for 3.5 hours of acquisition. After 3 hours, the air conditioning in the laboratory is switched off, which leads to the misalignment of the optical components. In turn, this leads to a significant drop of the witness value $W$, and in the entropy estimation. Hence a larger compression is required in the post-processing, resulting in a lower rate, but importantly high quality randomness is still guaranteed. (b) NIST tests of the data at the output of the extractor. For each test the represented p-value is the result of the Kolmogorov-Smirnov test. To pass the test the p-value needs to satisfy $0.01 \leq p \leq 0.99$. (c) Binary image of the extracted random bits. The image is a $500 \times 500$ matrix. No pattern appears.

cussed in App. D. The size of the randomness extractor is determined based on the value of $W$, which allows us to extract random bits [15, 29]. We use a Toeplitz-matrix extractor based on a pseudo-random seed. In the best conditions, our setup generates about 402 bits/s of raw data (before the extractor). The witness corresponds to a value of $W = 0.76$ (due to a visibility of 0.87 with the BB84 states and measurements). 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 $\sim 8\%$; total efficiency $\sim 2\%$). For a particular run of 3.5 hours, Fig. 3 shows the estimated value of $W$ and the rate at which the final random bits are generated. To demonstrate the self-testing capacity of our QRNG setup, we voluntarily stopped the air conditioning in the room after 3 hours. As can be clearly seen from Fig. 3(a), this event impacts BM by changing the preparation and measurement settings. As a result, the witness value $W$ drops, thus decreasing the randomness generation rate. Therefore, our setup can guarantee the generation of high quality random numbers, without continuously verifying the setup’s alignment and without the need of modeling the impact of the temperature increase on the optical-fiber setup.

The quality of the generated randomness can be assessed by checking for patterns and correlations in the extracted bits. We performed standard statistical test, 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 FIG. 3(b)). 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. A more visual approach to detecting patterns is illustrated in FIG. 3(c), where we display 250000 bits in a $500 \times 500$ matrix as a black-and-white image. Any repeated pattern or regular structure in the image would indicate correlations among the bits. No pattern appears.

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 non-malicious devices. Note however that this is not necessary strictly speaking, as our protocol is in principle robust to arbitrarily low detection efficiency [23]. Performing the data analysis without the fair-sampling assumption (in which case the inconclusive events are attributed the outcome
IV. CONCLUSION

We presented a protocol for self-testing QRNG, which allows for the real-time monitoring of the entropy of the raw data. In turn, this allows one to adapt 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 robust approach to QRNG, which can be viewed as intermediate between the standard (device-dependent) approach and the device-independent one.

Finally, we briefly compare our approach to the fully device-independent one [17, 18]. While the latter offers a stronger form of security (in particular assumptions (ii)-(iv) can be relaxed, hence offering robustness to side-channels and memory effects), its practical implementation is extremely challenging. Proof-of-principle experiments require state-of-the-art setups but could achieve only very low rates [13, 19]. On the other hand, 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.

Acknowledgements. We thank Antonio Acin, Stefano Pironio, Valerio Scarani, and Eric Woodhead for discussions, we thank Raphael Houlmann and Claudio Barreiro for technical support, and we thank Batelle and ID Quantique for providing the PPLN waveguide. The authors acknowledge financial support from the Swiss National Science Foundation (grant PP00P2_138917 and QSIIT), SEFRI (COST action MP1006) and the EU DIQIP.

[1] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, Journal of Modern Optics 47, 595 (2000).
[2] T. Jennewein, U. Aichler, G. Weihs, H. Weinfurter, and A. Zeilinger, Review of Scientific Instruments 71, 1675 (2000).
[3] J. F. Dynes, Z. L. Yuan, A. W. Sharpe, and A. J. Shields, Applied Physics Letters 93, 031109 (2008).
[4] M. Wahl, M. Leitgen, M. Berlins, T. Ruhlicke, H.-J. Rahn, and O. Benson, Applied Physics Letters 98, 171105 (2011).
[5] Y.-Q. Nie, H.-F. Zhang, Z. Zhang, J. Wang, X. Ma, J. Zhang, and J.-W. Pan, Applied Physics Letters 104, 051110 (2014).
[6] M. Stipcevic and B. M. Rogina, Review of Scientific Instruments 78, 045104 (2007).
[7] B. Qi, Y.-M. Chi, H.-K. Lo, and L. Qian, Opt. Lett. 35, 312 (2010).
[8] A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, OowadaSao, T. Kurashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis, Nat Photon 2, 728 (2008).
[9] C. Abellán, W. Amaya, M. Jofre, M. Curty, A. Acín, J. Capmany, V. Pruneri, and M. W. Mitchell, Opt. Express 22, 1645 (2014).
[10] C. Gabriel, C. Wittmann, D. Sych, R. Dong, W. Maurer, U. L. Andersen, C. Marquardt, and G. Leuchs, Nat Photon 4, 711 (2010).
[11] T. Symul, S. M. Assad, and P. K. Lam, Applied Physics Letters 98, 231103 (2011).
[12] B. Sanguinetti, A. Martin, H. Zbinden, and N. Gisin, Phys. Rev. X 4, 031056 (2014).
[13] N. Nisan and A. Ta-Shma, Journal of Computer and System Sciences 58, 148 (1999).
[14] Y. Dodis, D. Pointcheval, S. Ruhault, D. Vergniaud, and D. Wichs, Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security CCS ’13, 647 (2013).
[15] D. Frauchiger, R. Renner, and M. Troyer, arXiv e-print (2013), arXiv:1311.4547 [quant-ph].
[16] X. Ma, F. Xu, H. Xu, X. Tan, B. Qi, and H.-K. Lo, Phys. Rev. A 87, 062327 (2013).
[17] R. Colbeck, Ph.d. thesis, arxiv:0911.3814.
[18] 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 464, 1021 (2010).
[19] 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).
[20] 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).
[21] H.-W. Li, M. Pawlowski, 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. Acín, arXiv e-print (2012), arXiv:1210.1272 [quant-ph].
[23] G. Vallone, D. G. Marangon, M. Tomasin, and PaoloVillore, arXiv e-print (2014), arXiv:1401.7917 [quant-ph].
[24] J. Bowles, M. T. Quintino, and N. Brunner, Phys. Rev. Lett. 112, 140407 (2014).
[25] C. H. Bennett and G. Brassard, Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India,175 (1984).
[26] S. Tanzilli, A. Martin, F. Kaiser, M. De Micheli, O. Al-
Appendix A: Proof of randomness

We provide an upper bound on the min-entropy of the output using the dimension witness of Ref. [23]. Although the devices are assumed to be independent, each device features an internal source of randomness, represented by the variable $\lambda$ for Alice, and variable $\mu$ for Bob. We denote the distributions of these variables $q(\lambda) = q_\lambda$ and $r(\mu) = r_\mu$.

Our goal is to upper bound the guessing probability averaged over all inputs and values of the local random variables. For given inputs and $\lambda$, $\mu$, the guessing probability is

$$p_{xy|\lambda\mu}^g = \max_b p(b|x, y, \lambda, \mu). \tag{A1}$$

The average guessing probability $p^g$ is the average of $p_{xy\lambda\mu}^g$ over the distribution of inputs and local randomness. To proceed, however, we will first derive an upper bound on $p_{xy\lambda\mu}^g$, defined to be the average over the inputs only.

We consider the witness $W$ of [14]. We thus have four preparations, $x = 0, 1, 2, 3$ and two measurements $y = 0, 1$. Consider choices of preparations and measurements which are uniformly random (as explained in the main text, pseudorandomness is sufficient here), i.e. each combination $x, y$ occurs with probability 1/8. We have that

$$p_{\lambda\mu}^q = \frac{1}{8} \sum_{x,y} \max_b p(b|x, y, \lambda, \mu) \leq \frac{1}{2} \sum_x \max_y \max_b p(b|x, y, \lambda, \mu) \leq \frac{1 + \cos(\theta_{\mu}/2)}{2} \tag{A2}$$

where $\theta_{\mu}$ denotes the angle between Bob’s two measurements. The reasoning of the derivation here is as follows.

The best guessing probability averaged over inputs of Alice is bounded by the maximum over her inputs. This gives the first inequality and allows us to focus on the best possible state that Alice can send. Next, Bob has two measurements described by Bloch vectors $T_{0,1}$, and

$$p_{\lambda\mu}^g \leq \frac{1}{2} \left( 1 + \sqrt{1 - W_{\lambda\mu}^2} \right) = f(W_{\lambda\mu}). \tag{A4}$$

We note that the function $f$ is concave and decreasing. Next, we establish the following convexity property of the witness (in a slight abuse of notation, $W$ denotes the observed value of the witness when $\lambda$, $\mu$ are not known)

$$W_{\lambda\mu} \leq |T_{0,1}^\mu| \leq \sin \theta_{\mu} \tag{A3}$$

For maximally anti-commuting measurements, we get $W_{\lambda\mu} = 1$. Combining (A2) and (A3), we get

$$\theta_{\mu} \text{ is the angle between them. The best guessing probability averaged over his inputs is obtained by sending a state which lies in the middle between his measurements on the Bloch sphere (see Fig. 4). For such a state, the outcome probabilities for the two values of } b \text{ are } \cos^2(\theta_{\mu}/4), \text{ and } \sin^2(\theta_{\mu}/4). \text{ Choosing the larger value and using the double-angle formula, one arrives at the second inequality.}

Now we use the fact that a bound on the angle $\theta_{\mu}$ can be derived from the witness value for fixed local randomness $W_{\lambda\mu}$. One has that (see [23])

$$W_{\lambda\mu} \leq |T_{0,1}^\mu| \leq \sin \theta_{\mu} \tag{A3}$$

For maximally anti-commuting measurements, we get $W_{\lambda\mu} = 1$. Combining (A2) and (A3), we get

$$W_{\lambda\mu} \leq \sum_{\lambda, \mu} q_\lambda r_\mu W_{\lambda\mu}. \tag{A5}$$

To see that this holds, consider the entries of the matrix defining $W$. They are of the form $p(1|x, y) - p(1|x', y)$.
When the devices have internal randomness, we can write

\[
p(1|x, y) - p(1|x', y) = \sum_{\lambda, \mu} q_{\lambda\mu} \left( \text{Tr}[\rho_2^\lambda \Pi_{1|y}^\mu] - \text{Tr}[\rho_2^{\lambda'} \Pi_{1|y}^{\mu'}] \right)
\]

\[
= \sum_{\lambda, \mu} q_{\lambda\mu} S_{23}^\lambda \cdot T_y
\]

\[
= \left( \sum_{\lambda} q_{\lambda} S_{23}^\lambda \right) \cdot \left( \sum_{\mu} r_{\mu} T_y^{\mu} \right)
\]

\[
\equiv S_{23} \cdot T_y, \quad (A6)
\]

where \(\rho_2^\lambda\) are the states produced by Alice’s box, and \(\Pi_{1|y}^{\mu} = (1 + M_y^\mu)/2\) are the projection operators of Bob corresponding to outcome 1, \(T_y^{\mu}\) is the Bloch vector corresponding to \(M_y^\mu\) and \(S_{23}^\lambda\) is the difference of the Bloch vectors for \(\rho_2^\lambda\) and \(\rho_2^{\lambda'}\), (see [23]). Now, from [23] it follows that

\[
W = (S_{01} \times S_{23}) \cdot (T_0 \times T_1) \quad (A7)
\]

\[
= \sum_{\lambda, \lambda', \mu, \mu'} q_{\lambda\lambda'} r_{\mu} r_{\mu'} (S_{01}^\lambda \times S_{23}^{\lambda'}) \cdot (T_0^\mu \times T_1^{\mu'}) \quad (A8)
\]

\[
= \sum_{\lambda, \lambda', \mu, \mu'} q_{\lambda\lambda'} r_{\mu} r_{\mu'} |S_{01}^\lambda \times S_{23}^{\lambda'}| |T_0^\mu \times T_1^{\mu'}| \cos \phi_{\lambda, \lambda', \mu, \mu'}
\]

\[
\leq \sum_{\lambda, \lambda', \mu, \mu'} q_{\lambda\lambda'} r_{\mu} r_{\mu'} |S_{01}^\lambda \times S_{23}^{\lambda'}| |T_0^\mu \times T_1^{\mu'}| \cos \phi_{\lambda, \lambda', \mu, \mu'} \quad (A9)
\]

\[
\leq \sum_{\lambda, \lambda', \mu} q_{\lambda\lambda'} |S_{01}^\lambda \times S_{23}^{\lambda'}| |T_0^\mu \times T_1^{\mu'}| \cos \phi_{\lambda, \mu} \quad (A10)
\]

\[
= \sum_{\lambda, \mu} q_{\lambda\mu} W_{\lambda, \mu} \quad (A11)
\]

Here the \(\phi\)'s denote the angle between the two cross products in the preceding brackets. In the fourth line we have reasoned as follows: for fixed \(\lambda, \mu, \mu'\) there will be a value of \(\lambda'\) such that \(|S_{01}^\lambda \times S_{23}^{\lambda'}| \cos \phi_{\lambda, \lambda', \mu, \mu'}\) is maximal. If we label this value \(\lambda\) and set \(q_{\lambda\lambda'} = 1\) when \(\lambda' = \lambda\) this can only increase the expression. A similar argument eliminates \(\mu'\).

We are now ready to bound the guessing probability \(p^g\). Using the definition of \(p^g\), (A3) and (A5) we have

\[
p^g = \sum_{\lambda, \mu} q_{\lambda\mu} p_{\lambda, \mu}^g \quad (A13)
\]

\[
\leq \sum_{\lambda, \mu} q_{\lambda\mu} f(W_{\lambda, \mu}) \quad (A14)
\]

\[
f(\sum_{\lambda, \mu} q_{\lambda\mu} W_{\lambda, \mu}) \quad (A15)
\]

\[
f(W) \quad (A16)
\]

where in the third line we have used Jensen’s inequality and concavity of \(f\), and in the last line we have used that \(f\) is decreasing. Hence, we finally get

\[
p^g \leq \frac{1}{2} \left( 1 + \sqrt{1 - W^2} \right) \quad (A17)
\]

which gives the desired upper bound on the guessing probability as a function of the observed value of the witness \(W\). This bound is tight when maximal violation of the witness is achieved, i.e. \(W = 1\). In App. [D] we provide the calculation for the maximum number of extractable random bits.

Finally, we provide a proof of the relation between \(W\) and the commutativity of the measurements (Eq. (5) in the main text). We write \(M_y^\mu = T_y^{\mu} \cdot \vec{\sigma}\), and we have

\[
\int d\mu r(\mu) || [M_0^\mu, M_1^\mu] || = \int d\mu r(\mu) || [T_0^{\mu} \cdot \vec{\sigma}, T_1^{\mu} \cdot \vec{\sigma}] ||
\]

\[
= \int d\mu r(\mu) \| 2(\vec{T}_0^{\mu} \times \vec{T}_1^{\mu}) \cdot \vec{\sigma} \|
\]

\[
\geq 2 \int d\lambda d\eta(\lambda)r(\mu) W_{\lambda, \mu}
\]

\[
\geq 2W, \quad (A18)
\]

where we have used [A3] and [A5].

Appendix B: Certifying randomness in the presence of afterpulsing

In the following we show that although afterpulsing a priory violates the i.i.d. assumption (iii), the self-testing nature of our protocol captures the effect. When afterpulsing is present, the witness value is reduced correspondingly and randomness can still be certified.

To see this, we first consider a hypothetical experiment in which the outputs are generated as follows: in a fraction \(\eta\) of events, the experiment follows and ideal quantum qubit implementation while for the remaining events an outcome is generated at random by the measurement device, determined only by some internal random variable \(\mu\) independent of the inputs. Let us denote the witness value computed from the whole dataset \(W\), and the value which would be obtained from only the quantum events \(\bar{W}\). To an observer who does not know \(\mu\), the non-quantum events look just like uniform noise and the witness values fulfill \(W = \eta^2 \bar{W}\) [23]. At the same time, this scenario meets all of the assumptions in the proof of randomness of App. [A]. Therefore, for an observer with perfect knowledge of \(\mu\), who can hence perfectly predict the output for the non-quantum events, the guessing probability on the whole dataset is bounded by

\[
p^g \leq f(W) = f(\eta^2 \bar{W}). \quad (B1)
\]

We now show that the witness value is reduced in a similar way for afterpulsing, and hence even if the outputs
from afterpulsing events can be perfectly predicted, our bound on the randomness still holds.

Consider an experiment generating a set $S = \{(b_1, x_1, y_1), \ldots, (b_N, x_N, y_N)\}$ of $N$ events. The first thing to notice is that afterpulsing is probabilistic: in any given event either there is an afterpulse or there is not. We can therefore think of $S$ as consisting of a set $\tilde{S}$ of $N$ events with no afterpulse and $N - N$ additional afterpulsing events. Let $N_{bxy}$ denote the number of events in $S$ with outcome $b$ and inputs $x$, $y$, and $\tilde{N}_{bxy}$ the events in $\tilde{S}$, and define $N_{xy}$ and $\tilde{N}_{xy}$ similarly. For simplicity let us consider the limit of large $N$ such that finite size effects can be neglected. Since the inputs are chosen uniformly $N_{xy} = N/8$. We note that the probability for an afterpulse to occur in a given round $i$ of the experiment does not depend on the inputs $x_i$, $y_i$ in that round. The number of afterpulses is therefore the same for all combinations of $x,y$, and $\tilde{N}_{xy} = \eta N/8$ with $\eta = N/N$. In any afterpulsing event, the outcome $b_i$ is also uncorrelated to the inputs $x_i$, $y_i$ in that round (since $b_i = b_{i-1}$). This means that the effect of afterpulsing when counting events can be written

$$N_{b,xy} = \tilde{N}_{b,xy} + c_b,$$

where, importantly, $c_b$ is independent of $x$ (also of $y$ and indeed may be independent of $b$, but this is not important in the following).

The witness value on the dataset $S$ is computed from the frequencies $\nu_{b,xy} = N_{b,xy}/N_{xy}$. Using the above, we can write

$$\nu_{b,xy} = \frac{\tilde{N}_{b,xy} + c_b}{N/8} = \frac{\eta \tilde{N}_{b,xy}}{\eta N/8} + \frac{8c_b}{N} = \eta \tilde{\nu}_{b,xy} + \frac{8c_b}{N},$$

where $\tilde{\nu}_{b,xy} = \tilde{N}_{b,xy}/\tilde{N}_{xy}$ is the frequency one would have obtained considering only the set $\tilde{S}$. Now, since the last term above is independent of $x$ and since the witness is computed solely from terms of the form $\nu_{1,xy} - \nu_{1,x'y}$, we have that

$$W = \eta^2 \bar{W},$$

where $\bar{W}$ is the witness value which one would obtain from the events $\tilde{S}$ without afterpulsing. Since the reduction in $W$ when afterpulses are added is exactly the same as in the scenario above where events with perfectly predictable outputs were added, it follows that even if afterpulse events would be perfectly predictable, the bound (B1) on the guessing probability still holds.

### Appendix C: Accounting for multi-photon events

For real-world sources it is challenging to guarantee that they are of qubit nature. In particular, single-photon sources based on spontaneous parametric down conversion process or weak coherent sources have non-zero probability of emitting more than one photon, violating the qubit assumption.

Given an imperfect source which does not always satisfy the qubit assumption, we would like to say something about the witness violation corresponding to events that do satisfy the assumption. In particular, we would like a lower bound on this violation in terms of the observed, experimental probability distribution and some guarantee on the fraction of non-qubit events. Even without a detailed model of the source, it is possible to determine this fraction e.g. using knowledge of the photon statistics.

### 1. Bounding the violation for given qubit fraction

To derive a bound on the quantum violation, we will assume that each experimental run either satisfies the qubit assumption, or not. That is, the conditional probability distribution for the experiment can be modeled as

$$p(b|zx) = \alpha p_{qa}(b|xz) + (1 - \alpha)p_{qa}(b|xz),$$

where $\alpha$ is the fraction of qubit events, $p_{qa}$ is the distribution corresponding to the qubit events, and $p_{qa}$ is an unrestricted distribution. The witness value is given in terms of the probabilities by $|W|$, where

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

From the model (C1), it follows that the expected witness value must satisfy

$$W = |\alpha^2 W_{qa} + (1 - \alpha)^2 W_{qa} + \alpha(1 - \alpha)\eta(G + G')|, \quad \text{(C3)}$$

where $W_{qa}$, $W_{qa}$ are the determinants corresponding to distributions $p_{qa}$ and $p_{qa}$ respectively, and

$$G = \begin{vmatrix} p_{qa}(1|0,0) - p_{qa}(1|1,0) & p_{qa}(1|2,0) - p_{qa}(1|3,0) \\ p_{qa}(1|0,1) - p_{qa}(1|1,1) & p_{qa}(1|2,1) - p_{qa}(1|3,1) \end{vmatrix},$$

$$G' = \begin{vmatrix} p_{qa}(1|0,0) - p_{qa}(1|1,0) & p_{qa}(1|2,0) - p_{qa}(1|3,0) \\ p_{qa}(1|0,1) - p_{qa}(1|1,1) & p_{qa}(1|2,1) - p_{qa}(1|3,1) \end{vmatrix}.$$

To bound the qubit violation for a given observed experimental violation we should minimise $|W_{qa}|$ subject to the constraint (C3). However, if a certain value $W$ can be attained for a fixed value of $|W_{qa}|$, then attaining all smaller values requires even less qubit violation. We may therefore just as well look for the maximal $W$ for fixed $|W_{qa}|$. Any value above this maximum guarantees a qubit violation of at least $|W_{qa}|$. The maximum has a simple form. It is given by

$$\max W = \max \left\{ 4\alpha(1 - \alpha) + \alpha(2\alpha - 1)W_{qa} \right\}. \quad \text{(C4)}$$

The first thing we notice is that when $\max W$ in (C4) is less than 1, it is always given by the first line. This is the
relevant case for certifying randomness in practice. Solving for the qubit violation, given an observed violation less than unity we have the bound

$$W_{qα} ≥ \frac{1}{α(2α - 1)}[W - 4α(1 - α)].$$  \hspace{1cm} (C5)$$

Second, we note that for $α > 1/2$ the maximum (C4) is always larger than 1. This means that to be able to certify randomness in practice, we need a minimal fraction of events satisfying the qubit assumption of

$$α > \frac{1}{2}.\hspace{1cm} (C6)$$

Third, for a given value of $α$ there is a minimal observed violation below which the bound (C5) becomes trivial and no randomness can be certified. We must have

$$W > 4α(1 - α).$$  \hspace{1cm} (C7)$$

2. Estimating the qubit fraction

For an implementation with a particular source, we need an estimate or a lower bound on the fraction of qubit events $α$. Source and detector inefficiency, and transmission losses lead to inconclusive events, and our estimate of $α$ should be consistent with how these events are dealt with.

In the scenario of non-malicious, error-prone devices considered here, it is rather natural to discard inconclusive events (e.g. assuming fair-sampling) and then compute $W$ from the remaining data. To be able to evaluate (C5) in this case, one needs to estimate $α$ when inconclusive events are discarded. It is also natural to assume that all events with at most one photon emitted obey the qubit assumption.

With these assumptions, let $q$ denote the probability for the source to emit at most one photon and consider an experiment with $N$ events and $M$ conclusive events. Before post-selection, asymptotically the fraction of events that obey the qubit assumption is then $α = q$. For a finite number of events, we can put a conservative estimate, i.e., a lower bound, on the number of events $N_α$ that satisfy the qubit assumption, within a given confidence. In particular, under the assumption that we know $q$, the behaviour of the source is modelled by a family of $N$ Bernoulli trials parameterized by $q$, and thus the estimation problem can be solved by using the Chernoff-Hoeffding tail inequality. More formally, let $ν > 0$ be the failure probability of the estimation process and $t > 0$ be the margin parameter, then

$$P(N_α ≤ qN - t) ≤ \exp(-2Nt^2) = ν, \hspace{1cm} (C8)$$

which implies that $N_α > qN - t$ is true with probability at least $1 - ν$. Equivalently, the fraction of qubit events without post-selection is $α > q - t/N$ with probability at least $1 - ν$. The margin parameter $t$ can be expressed in terms of $N$ and $ν$ as $t = \sqrt{1/(2N)\log(1/ν)}$.

To account for post-selection, we conservatively assume that all multi-photon events are conclusive. Asymptotically, the fraction of non-qubit events will be $(1 - q)N/M$, so $α = 1 - (1 - q)N/M$. For finite $N$ we have that after post-selection

$$α ≥ 1 - (1 - q)\frac{N}{M} - \frac{t}{M}$$  \hspace{1cm} (C9)$$

with probability at least $1 - ν$, with $ν$ and $t$ given by (C8).

Appendix D: Security Analysis

In this section, we show that with the observed experimental statistics, it is possible to provide a bound on the number of random bits that can be extracted from the raw data set, $Z$, which takes values from a set of all binary strings, $Z$ of length $m$. Our approach essentially uses the (quantum) leftover hash lemma, which states that the amount of private randomness is approximately equal to the min-entropy characterization of the raw data $Z$. More specifically, it says that the number of extractable random bits (that is independent of variables $X,Y,L$) is roughly given by $H_{min}(Z|XYL)$. Here, we recall that variables $X$ and $Y$ are the inputs of Alice and Bob, respectively, and $L$ is the classical register capturing all information about the local variables $λ$ and $µ$. The min-entropy of $Z$ given $XYL$ has a clear operational meaning when casted in terms of the guessing probability, i.e., $H_{min}(Z|XYL) = -m \log_2 p_{guess}$: it measures the probability of correctly guessing $Z$ when given access to classical side-information $XYL$.

On a more concrete level, the leftover hash lemma employs a family of universal hash functions to convert $Z$ into an output string $S$ (of size $ℓ$) that is close to a uniform string conditioned on side-information $XYL$. In particular, we say that the output string $S$ is $Δ$-close to uniform conditioned on $XYL$, if

$$\frac{1}{2} \sum_{s.x.y.l} |P_{sXYL} - U_{SPXYL}| ≤ Δ,$$  \hspace{1cm} (D1)$$

where $U_δ$ is the uniform distribution of $S$. The quality of the output string is directly related to the number of extractable random bits, i.e.,

$$ℓ = \left[ H_{min}(Z|XYL) - 2\log_2 \frac{1}{2Δ} \right].$$  \hspace{1cm} (D2)$$

Therefore, to bound $ℓ$, we only need to fix a security level $ε_{sec} ≥ Δ$ and find a lower bound on the min-entropy term. Using the definition of conditional min-entropy and the assumption that $Z$ is generated from an iid process, we
have
\[ \ell = m - m \log_2 \left( 1 + \sqrt{1 + \frac{1 - W^2}{2}} - 2 \log_2 \frac{1}{2\Delta} \right). \]  
(D3)

Accordingly, the rate of extraction is \( \ell/m \) and it converges to the min-entropy rate when \( m \to \infty \) (therefore \( \Delta \to 0 \)). At the moment, our bound on \( \ell \) is written in terms of the expected value of \( W \), which is not directly accessible in the experiment. In order to relate the \( W \) to the set of experimental statistics \( \mathcal{E} := \{ n_{x,y}^+ / n_{x,y} \}_{x,y} \), we first use the Chernoff-Hoeffding tail inequality to get
\[ p(1|x,y) - t(\epsilon_{pe}, n_{x,y}) \leq \frac{\epsilon_{pe} n_{x,y}^+}{n_{x,y}} \leq p(1|xy) + t(\epsilon_{pe}, n_{x,y}), \]  
(D4)
where \( t(\epsilon_{pe}, n_{x,y}) := \sqrt{\log(1/\epsilon_{pe})/(2n_{x,y})} \). Here, relations with oversetting \( \epsilon_{pe} \) means that the relations are probabilistically true, i.e., the relations hold except with probability \( \epsilon_{pe} \). For our purposes later, we denote \( p_{x,y}^\pm := p(1|x,y) \pm t(\epsilon_{pe}, n_{x,y}) \). In the following, we introduce an estimate of the expected \( W \), i.e.,
\[ W \geq W_{\min} := \min_{q_{x,y} \in \{ p_{x,y}, p_{x,y}^\pm \}} |W(q_{x,y})|, \]  
(D5)
where \( \epsilon' = 8\epsilon_{pe} \) and
\[ W(q_{x,y}) := \det \begin{bmatrix} q_{0,0} - q_{1,0} & q_{2,0} - q_{3,0} \\ q_{0,1} - q_{1,1} & q_{2,1} - q_{3,1} \end{bmatrix}. \]  
(D6)

Next, we need to bound the maximum fraction of non qubit events, \( 1 - \alpha \). Following the discussion in App. C with post-selection we expect \( \alpha \) to be \( 1 - \frac{p_2}{p_1 + p_2} \) (\( p_2 \) and \( p_1 \) are the probabilities of the SPDC to emit, respectively, a double pair or a single-photon pair). In the scenario where \( N \) preparations are made, by using the Chernoff-Hoeffding tail inequality, we have that
\[ \alpha' \geq \hat{\alpha} := 1 - \frac{p_2}{\epsilon''} + t(\epsilon'', N). \]  
(D7)

Plugging this into Eq. (C5), we get
\[ W_{qa} \geq W_{\min} - 4\hat{\alpha}(1 - \hat{\alpha}) (2\hat{\alpha} - 1). \]  
(D8)

Therefore, the effective violation is
\[ \hat{W}_{\text{eff}} := \frac{W_{\min} - 4\hat{\alpha}(1 - \hat{\alpha})}{2\hat{\alpha} - 1}. \]  
(D9)

Note that the effective violation is obtained by fixing the violation due to non qubit contribution to be zero. In other words, the effective violation measures the amount of randomness in \( Z \). That is, we have
\[ \ell = m - m \log_2 \left( 1 + \sqrt{1 + \frac{1 - \hat{W}_{\text{eff}}^2}{2}} - 2 \log_2 \frac{1}{2\Delta} \right). \]
Finally, by choosing \( \Delta = \epsilon \) and fixing \( \epsilon_{pe} = \epsilon'' = \epsilon \), the output string \( S \) is \( 10\epsilon \)-close to uniform conditioned on \( XYL \). In the actual implementation we chose \( \epsilon = 10^{-3} \).