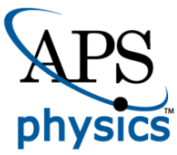

This is the accepted manuscript made available via CHORUS. The article has been published as:

Experimental Low-Latency Device-Independent Quantum Randomness

Yanbao Zhang, Lynden K. Shalm, Joshua C. Bienfang, Martin J. Stevens, Michael D. Mazurek, Sae Woo Nam, Carlos Abellán, Waldimar Amaya, Morgan W. Mitchell, Honghao Fu, Carl A. Miller, Alan Mink, and Emanuel Knill

Phys. Rev. Lett. **124**, 010505 — Published 7 January 2020

DOI: 10.1103/PhysRevLett.124.010505

Experimental Low-Latency Device-Independent Quantum Randomness

Yanbao Zhang,[1,*] Lynden K. Shalm,[2] Joshua C. Bienfang,[3] Martin J. Stevens,[2] Michael D. Mazurek,[2] Sae Woo Nam,[2] Carlos Abellán,[4,†] Waldimar Amaya,[4,†] Morgan W. Mitchell,[4,5] Honghao Fu,[6] Carl A. Miller,[6,3] Alan Mink,[3,7] and Emanuel Knill[2,8]

[1] *NTT Basic Research Laboratories and NTT Research Center for Theoretical Quantum Physics, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan*
[2] *National Institute of Standards and Technology, Boulder, Colorado 80305, USA*
[3] *National Institute of Standards and Technology, Gaithersbug, MD 20899, USA*
[4] *ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[5] *ICREA-Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain*
[6] *Department of Computer Science, Institute for Advanced Computer Studies, and Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, MD 20742, USA*
[7] *Theiss Research, La Jolla, CA 92037, USA*
[8] *Center for Theory of Quantum Matter, University of Colorado, Boulder, Colorado 80309, USA*

Applications of randomness such as private key generation and public randomness beacons require small blocks of certified random bits on demand. Device-independent quantum random number generators can produce such random bits, but existing quantum-proof protocols and loophole-free implementations suffer from high latency, requiring many hours to produce any random bits. We demonstrate device-independent quantum randomness generation from a loophole-free Bell test with a more efficient quantum-proof protocol, obtaining multiple blocks of 512 random bits with an average experiment time of less than 5 min per block and with certified error bounded by $2^{-64} \approx 5.42 \times 10^{-20}$.

A fundamental feature of quantum mechanics is that measurements of a quantum system can have random outcomes even when the system is in a definite, pure state. By definition, pure states are completely uncorrelated with every other physical system, which implies that the measurement outcomes are intrinsically unpredictable by anyone outside the measured quantum system's laboratory. The unpredictability of quantum measurements is exploited by conventional quantum random number generators (QRNGs) [1] for obtaining random bits whose distribution is ideally uniform and independent of other systems. The use of such QRNGs requires trust in the underlying quantum devices [2]. A higher level of security is attained by device-independent quantum random number generators (DIQRNGs) [3, 4] based on loophole-free Bell tests, where the randomness produced can be certified even with untrusted quantum devices that may have been manufactured by dishonest parties. The security of a DIQRNG relies on the physical security of the laboratory to prevent unwanted information leakage, and on the trust in the classical systems that record and process the outputs of quantum devices for randomness generation.

Since the idea of DIQRNGs was introduced in Colbeck's thesis [3], many DIQRNG protocols have been developed—for a review see [5]. These protocols generally exploit quantum non-locality to certify entropy but differ in device requirements, Bell-test configurations, randomness rates, finite-data efficiencies, and the security levels achieved. We can classify protocols by whether they are secure in the presence of classical or quantum side information, in other words, by whether they are classical- or quantum-proof.

The first experimentally accessible DIQRNG protocol was given and implemented by Pironio *et al.* [6] with a detection-loophole-free Bell test using entangled ions. They certified 42 bits of classical-proof entropy with error bounded by 0.01, where, informally, the error can be thought of as the probability that the protocol output does not satisfy the certified claim. This required about one month of experiment time. To improve this result required the advent of loophole-free Bell tests and much more efficient protocols. Such a protocol and experimental implementation with an optical loophole-free Bell test was given by Bierhorst *et al.* [7] and obtained 1024 classical-proof random bits with error 10^{-12} in 10 min. There have been three demonstrations of quantum-proof DIQRNGs, all with photons. The first two were subject to the locality and freedom-of-choice loopholes [8]. They obtained 4.6×10^7 random bits with error 10^{-5} in 111 h [9], and 6.2×10^5 random bits with error 10^{-10} in 43 min [10], respectively. The third was loophole-free and obtained 6.2×10^7 random bits with error 10^{-5} in 96 h [11].

The quantum-proof experiments described above aimed for good asymptotic rates. To approach the asymptotic rate requires a very large number of trials to certify a large amount of entropy. However, many if not most applications of certified randomness require only short blocks of fresh randomness. To address these applications, we consider instead a standardized request for 512 random bits with error $2^{-64} \approx 5.42 \times 10^{-20}$ and with minimum delay, or latency, between the request and delivery of bits satisfying the request. In this

work, we consider only the contribution of experiment time to latency. The previous quantum-proof DIQRNG implemented with a loophole-free Bell test [11] would have required at least 24.1 h to satisfy the standardized request—see the Supplemental Material (SM) [12].

In this letter, we reduce the latency required to produce 512 device-independent and quantum-proof random bits with error 2^{-64} by orders of magnitude. For this purpose, here we implement a quantum-proof protocol developed in the companion paper (CP) [13] with a loophole-free Bell test. Unlike other demonstrations of quantum-proof DIQRNGs, we conservatively account for adversarial bias in the setting choices, and we show repeated fulfillment of the standardized request. We obtain five successive blocks of 512 random bits with error 2^{-64} and with an average experiment time of less than 5 min per block.

Overview of theory. We give a high-level description of the features of our protocol. For formal definitions and technical details, see the CP [13]. Our protocol is based on repeated (but not necessarily independent or identical) trials of a loophole-free CHSH Bell test [14], consisting of a source S and two measurement stations A and B (see Fig. 2). In each trial, the source attempts to distribute an entangled pair of photons to the stations, the protocol randomly chooses binary measurement settings X and Y for the stations, the corresponding measurements are performed, and the binary outcomes A and B are recorded. We call $Z = XY$ and $C = AB$ the input and output of the trial, respectively.

An end-to-end randomness generation protocol starts with a request for k random bits with error ϵ. The user then chooses a positive quantity σ (the entropy threshold for success) and positive errors $\epsilon_\sigma, \epsilon_x$ (the entropy error and the extractor error, respectively) whose sum is no more than ϵ. The quantity σ chosen by the user must satisfy the inequality $\sigma \geq k + 4\log_2(k) + 4\log_2(2/\epsilon_x^2) + 6$. This inequality is sufficient to guarantee that, if the outputs of the experiment can be proven to have entropy at least σ, then k uniformly random bits can be extracted. (The randomness extractor that we use for this purpose is Trevisan's extractor [15] as implemented by Mauerer, Portmann and Scholz [16]. We refer to it as the TMPS extractor—see the SM [12].) The user also needs to decide the maximum number n of Bell-test trials to run. For simplicity, we temporarily assume that a fixed number n of trials will be executed, but in the implementation as described in a later section we exploit the ability to stop early.

After fixing the parameters defined in the previous paragraph, n Bell-test trials are sequentially executed, and the inputs and outputs are recorded as $\mathbf{Z} = (Z_i)_{i=1}^n$ and $\mathbf{C} = (C_i)_{i=1}^n$, where Z_i and C_i are the input and output of the i'th trial. The upper-case symbols $\mathbf{C}$, C_i, $\mathbf{Z}$ and Z_i are treated as random variables, and their values are denoted by the corresponding lower-case symbols. Let E denote the "environment" of the experiment, including any quantum side information that could be possessed by an adversary. The entropy of the outputs $\mathbf{C}$ is quantified by the quantum ϵ_σ-smooth conditional min-entropy of $\mathbf{C}$ given $\mathbf{Z}\mathsf{E}$ [17]. We refer to this quantity as the output entropy. The user can estimate the output entropy as described in the next section and check whether that estimate is at least σ. If not, the protocol fails and a binary variable P is set to $P = 0$; otherwise, the protocol succeeds and $P = 1$.

When the protocol succeeds, we apply the TMPS extractor [16] to extract k random bits with error ϵ. The TMPS extractor is a classical algorithm that is applied to the outputs $\mathbf{C}$ as well as a random seed S, and produces a bit string R. The final state of the protocol then consists of the classical variables $RS\mathbf{Z}P$ and the quantum system E. In the CP [13], we prove that the protocol is ϵ-sound in the following sense: The error ϵ is an upper bound on the product of the success probability and the purified distance [18] between the actual state of $RS\mathbf{Z}\mathsf{E}$ conditional on the success event $P = 1$ and an ideal state of $RS\mathbf{Z}\mathsf{E}$, according to which RS is uniformly random and independent of $\mathbf{Z}\mathsf{E}$. For the protocol to be useful, it is necessary that the probability of success in the actual implementation can be close to 1, a property referred to as completeness. With properly configured quantum devices, it is possible to make this probability exponentially close to 1 by increasing the number of trials executed. Soundness and completeness imply formal security of the protocol.

Estimating entropy. In the CP [13], we develop the approach of certifying entropy by "quantum estimation factors" (QEFs), a general technique that encompasses previous certification techniques against quantum side information [19, 20]. The construction of QEFs requires first defining a notion of models. The "model" for an experiment is the set of all possible final states that can occur at the end of the experiment. A final state can be written as $\rho_{\mathbf{CZ}\mathsf{E}} = \sum_{\mathbf{cz}} |\mathbf{cz}\rangle\langle\mathbf{cz}| \otimes \rho_{\mathsf{E}}(\mathbf{cz})$, where $\rho_{\mathsf{E}}(\mathbf{cz})$ is the unnormalized state of E given results $\mathbf{cz}$.

Given the state $\rho_{\mathbf{CZ}\mathsf{E}}$, we characterize the unpredictability of the outputs $\mathbf{c}$ given the system E and the inputs $\mathbf{z}$ by the sandwiched Rényi power, denoted by $\mathcal{R}_{1+\beta}\big(\rho_{\mathsf{E}}(\mathbf{cz})\big|\rho_{\mathsf{E}}(\mathbf{z})\big)$ where $\beta > 0$ and $\rho_{\mathsf{E}}(\mathbf{z}) = \sum_{\mathbf{c}} \rho_{\mathsf{E}}(\mathbf{cz})$ (see the SM [12] for the explicit expression). A QEF with a positive power β for a sequence of n trials is a non-negative function T of random variables $\mathbf{CZ}$ such that for all states $\rho_{\mathbf{CZ}\mathsf{E}}$ in the model, T satisfies the inequality

$$\sum_{\mathbf{cz}} T(\mathbf{cz}) \mathcal{R}_{1+\beta}\left(\rho_{\mathsf{E}}(\mathbf{cz})|\rho_{\mathsf{E}}(\mathbf{z})\right) \leq 1.$$

Informally, one main result in the CP [13] is that if at the conclusion of the experiment the variable $\log_2(T)/\beta$ takes a value at least h for some $h > 0$, then the output

entropy (in bits) must be at least $h - \log_2(2/\epsilon_\sigma^2)/\beta$ no matter which particular state in the model describes the experiment. Hence, for estimating entropy it suffices to construct QEFs.

In practice, the model for a sequence of trials is constructed as a chain of models for each individual trial. QEFs then satisfy a chaining property: If $F_i(C_i Z_i)$ is a QEF with power β for the i'th trial, then the product $\prod_{i=1}^n F_i(C_i Z_i)$ is a QEF with power β for the sequence of n trials. To construct the QEF $T(\mathbf{CZ})$, we use this property. Moreover, since the model for each trial of our experiment is identical, we always take the same QEF for each executed trial. The CP [13] contains general techniques for constructing models and QEFs, and the SM [12] contains the details of constructing models and QEFs for each trial of our experiment.

Experiment. Our setup is similar to those reported in Refs. [7, 21]. A pair of polarization-entangled photons are generated through the process of spontaneous parametric downconversion and then distributed via optical fiber to Alice and Bob (see Fig. 1). At each lab of Alice and Bob, a fast QRNG with parity-bit randomness extraction [22] is used to randomly switch a Pockels cell-based polarization analyzer (see Fig. 2). Alice's polarization measurement angles, relative to a vertical polarizer, are $a = 4.1°$ and $a' = 25.5°$, and Bob's are $b = -a$ and $b' = -a'$. These measurement angles, along with the non-maximally entangled state prepared in Fig. 1, are chosen based on numerical simulations of our setup to achieve an optimal Bell violation. The photons are then detected in each lab using superconducting nanowire single-photon detectors with efficiency greater than 90% [23]. The total system efficiencies for Alice and Bob are $76.2 \pm 0.3\%$ and $75.8 \pm 0.3\%$, allowing the detection loophole to be closed. With the configuration detailed in Fig. 2, we can also close the locality loophole.

In each trial, Alice's and Bob's setting choices X and Y are made with random bits whose deviation from uniform is assumed to be bounded. That is, knowing all events in the past light cone, one should not be able to predict the next choice with a probability better than $0.5 + \epsilon_b$. We call ϵ_b the (maximum) adversarial bias. In particular, it is assumed that the quantum devices used cannot have more prior knowledge of the random setting choices than the adversarial bias for each trial. Specifically, we assume that the adversarial and trial-dependent bias of Alice's and Bob's QRNGs is bounded by $\epsilon_b \leq 1 \times 10^{-3}$. That is, each of the setting choices X and Y has a two-outcome distribution with probabilities in the interval $[0.5 - 1 \times 10^{-3}, 0.5 + 1 \times 10^{-3}]$. The bias assumption is supported in two ways: first by a quantum statistical model of the QRNGs, validated by measurements of the QRNG internal operation [22], and second by the observation that the frequencies of the output bits of each QRNG deviate from 0.5 by less than 6×10^{-5} on average in a run of 21 min of trials.

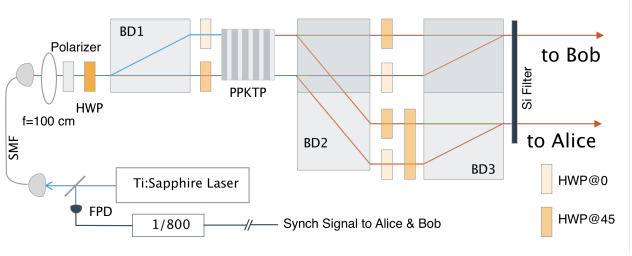

FIG. 1. Diagram of the entangled photon-pair source. A 775-nm-wavelength picosecond Ti:Sapphire laser operating at a 79.3 MHz repetition rate pumps a 20-mm-long periodically-poled potassium titanyl phosphate (PPKTP) crystal, to produce degenerate photons at 1550 nm with a per-pulse probability of 0.0045. The pump is transmitted through a polarization-maintaining single-mode fiber (SMF). The PPKTP crystal is cut for type-II phasematching and placed in a polarization-based Mach-Zehnder interferometer constructed using half-wave plates (HWPs) and three beam displacers (BD1, BD2 and BD3). Tuning the polarization of the pump by a polarizer and HWP allows us to create the non-maximally entangled state $|\psi\rangle = 0.967\,|HH\rangle + 0.254\,|VV\rangle$, where H and V denote the horizontally and vertically polarized single-photon states. The photons, along with a synchronization signal, are then distributed via optical fiber to Alice and Bob. The synchronization signal is generated by a fast photodiode (FPD) and divider circuit which devides the pump frequency by 800, and is used as a clock to determine the start of a trial and to time the operation of Alice's and Bob's measurements. This leads to a trial rate of approximately 100 kHz.

Protocol implementation. The goal is to obtain $k = 512$ random bits with error $\epsilon = 2^{-64}$. For this, we set $\epsilon_\sigma = 0.8 \times 2^{-64}$ and $\epsilon_x = 0.2 \times 2^{-64}$. To extract $k = 512$ random bits with the TMPS extractor, it suffices to set the entropy threshold to be $\sigma = 1089$. The implementation stages for each instance of the protocol are summarized in Box 1, and more details are available in the SM [12].

Results. Ideally, the protocol would be applied concurrently with the acquisition of the experimental trials. In this case, the trials were performed three months before the protocol was fully implemented. About 89 min of experimental results were recorded. The results were stored in 1 min blocks containing approximately 6×10^6 trials each. The first 21 min were unblinded for testing the protocol, and the rest were kept in blind storage until the protocol was fully implemented and ready to be used.

From the first 21 min of unblinded results we decided to run five sequential instances of the protocol, and for calibration in each instance we determined to use the 10 min of results preceding to the first trial to be used for ranodomness accumulation (see the SM [12] for details). We note that the trials for randomness accumulation in one instance can be used also for calibration in the next instance. For the protocol, we loaded the data

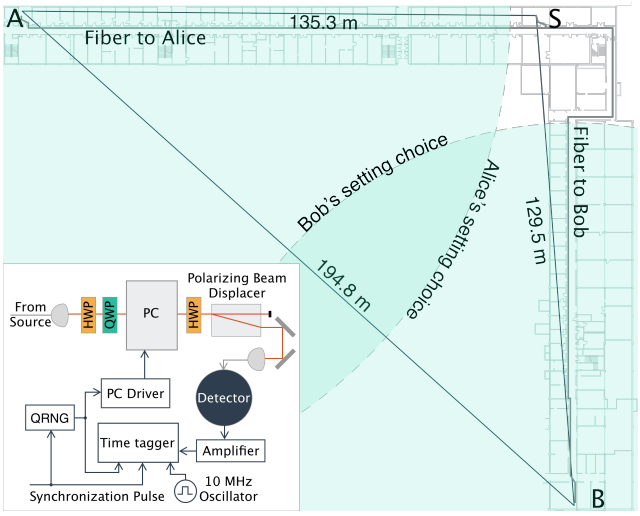

FIG. 2. Locations of Alice (A), Bob (B), and the source (S). Alice and Bob are separated by 194.8 ± 1.0 m (this is slightly further than in Refs. [7, 21]). Faint grey lines indicate the paths that the entangled photons take from the source to Alice and Bob through fiber optic cables. The light-green quarter circles are the 2D projections of the expanding light spheres containing the earliest available information about the random bits used for Alice's and Bob's setting choices at the trial. When Bob finishes his measurement, the radius of the light sphere corresponding to the start of Alice's QRNG has expanded to 127.3 ± 0.5 m, after which it takes an additional 222.3 ± 3.8 ns before the light sphere will intersect Bob's location. Similarly, when Alice completes her measurement, the light sphere corresponding to the start of Bob's QRNG has only reached a radius of 98.3 ± 0.5 m, and it will take 315.5 ± 3.8 ns more to arrive at Alice's station. In this way, the actions of Alice and Bob are spacelike separated. Inset: Alice's and Bob's measurement apparatuses both consist of a Pockels cell (PC), operating at approximately 100 KHz, and a polarizer, constructed using two have-wave plates (HWPs), a quarter-wave plate (QWP) and a polarizing beam displacer, in order to make fast polarization measurements on their respective photons. The measurement setting is controlled by a QRNG, the photon is detected by a high-efficiency superconducting nanowire single-photon detector, and the resulting signal is recorded on a time tagger, where a 10 MHz oscillator is used to keep Alice's and Bob's time taggers synchronized.

Box 1: Overview of protocol implementation

1. Calibration
   (a) Determine the QEF $F(CZ)$ and its power β used for each executed trial.
   (b) Fix n—the maximum number of trials.
2. Randomness Accumulation: Run the experiment to acquire up to n trials. After each trial i,
   (a) Update the running $\log_2$-QEF value $L_i = \sum_{j=1}^{i} \log_2(F(c_j z_j))$, where c_j and z_j are the observed values of C_j and Z_j.
   (b) If $\left(L_i - \log_2(2/\epsilon_\sigma^2)\right)/\beta \geq \sigma$, stop the experiment, set the number of trials actually executed as $n_{\text{act}} = i$, and set the success event $P = 1$.
3. Randomness Extraction: If $P = 1$, then extract k random bits with error ϵ.

TABLE I. Characteristics of the five protocol instances. The number of subblocks is approximately the number of seconds of experiment time required. The entropy rate is estimated by $L_{n_{\text{act}}}/(\beta n_{\text{act}})$, where n_{act} is the actual number of trials executed in an instance, $L_{n_{\text{act}}}$ is the running $\log_2$-QEF value at the end of an instance, and β is the power associated with the QEF which is used for each executed trial and determined at the calibration stage. The trial rate in the experiment was approximately 100 kHz.

Instance	$n/10^7$	$n_{\text{act}}/10^7$	Number of sub-blocks	β	Entropy rate/10^{-4}
1	5.25	2.32	233	0.010	6.07
2	4.74	3.76	379	0.010	3.78
3	5.92	2.85	287	0.009	5.47
4	6.20	2.83	285	0.009	5.53
5	5.49	2.72	274	0.010	5.20

and divided each 1 min block into 60 subblocks of approximately 1×10^5 trials each. The protocol was then designed to use integer multiples of these subblocks. The first instance of the protocol started producing randomness at the 22nd 1 min block. Each instance started at the first not-yet-used subblock and used the previous 600 subblocks for calibration, then processed subblocks until the running entropy estimate surpassed the threshold σ. In each instance, this happened well before the maximum number of trials n determined at the calibration stage was reached, leading to success of the instance. We then applied the extractor to produce 512 random bits with error 2^{-64}.

The results are summarized in Tab. I. It shows that the experiment time required to fulfill the request for 512 quantum-proof random bits with error 2^{-64} is less than 5 min on average, demonstrating a dramatic improvement over other quantum-proof protocols and previous experiments. The only experimentally accessible alternative quantum-proof protocol is entropy accumulation as described in Ref. [20]. We found that satisfying the request using theoretical results from Ref. [20], with our experimental configuration and performance, would have required at least 6.108×10^{10} trials, corresponding to 169.7 h of experiment time—see the SM [12] for details.

In conclusion, we demonstrated five sequential in-

stances of the DIQRNG protocol. For joint (or composable) security of the five instances, it suffices that the quantum devices do not retain memory of what happened during the previous instances. Without this assumption, the joint security of the five instances can be compromised as explained in Ref. [24]. In our implementation such problems are mitigated by the definition of soundness in terms of the purified distance rather than the conventional trace distance, but the issues arising in composing protocols like ours need further investigation.

We have emphasized the importance of latency. To produce a fixed block of random bits, latency is simply the time it takes for the protocol to fulfill the request. Above, we have neglected the classical computing time required for calibration and extraction since this can be made relatively small by using faster and more parallel computers. For the current implementation the time costs for calibration and extraction are detailed in the SM [12]. The latency for our setup is limited by the rate at which we can implement random setting choices, which in turn is limited by the Pockels cells. Since the source produces pulses at a rate of 79.3 MHz and we can use 10 successive laser pulses as a single trial without reducing the quality of trials, if the Pockels cell limitation can be overcome, the latency could be reduced by a factor of about 80 with the current entangled photon-pair source.

This work includes contributions of the National Institute of Standards and Technology, which are not subject to U.S. copyright. The use of trade names does not imply endorsement by the U.S. government. The work is supported by the National Science Foundation RAISE-TAQS (Award 1839223); the European Research Council (ERC) projects AQUMET (280169), ERIDIAN (713682); European Union projects QUIC (Grant Agreement no. 641122) and FET Innovation Launchpad UVALITH (800901); the Spanish MINECO projects MAQRO (Ref. FIS2015-68039-P), the Severo Ochoa programme (SEV-2015-0522); Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) project (2014-SGR-1295); Fundació Privada Cellex and Generalitat de Catalunya (CERCA Program).

Y. Z. and L. K. S. contributed equally to this work.

* Corresponding author.
yanbaoz@gmail.com
† Current address: Quside Technologies S.L., C/Esteve Terradas 1, Of. 217, 08860 Castelldefels (Barcelona), Spain

[1] M. Herrero-Collantes and J. C. Garcia-Escartin. Quantum random number generators. *Rev. Mod. Phys.*, 89:015004, 2017.
[2] S. Pironio and S. Massar. Security of practical private randomness generation. *Phys. Rev. A*, 87:012336, 2013.
[3] R. Colbeck. *Quantum and Relativistic Protocols for Secure Multi-Party Computation*. PhD thesis, Trinity College, University of Cambridge, Cambridge, UK, 2006. arXiv:0911.3814.
[4] R. Colbeck and A. Kent. Private randomness expansion with untrusted devices. *J. Phys. A*, 44:095305, 2011.
[5] A. Acín and L. Masanes. Certified randomness in quantum physics. *Nature*, 540:213–219, 2016.
[6] S. Pironio, A. Acin, 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. 464:1021–1024, 2010.
[7] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, , M. J. Stevens, and L. K. Shalm. Experimentally generated random numbers certified by the impossibility of superluminal signaling. *Nature*, 556:223–226, 2018.
[8] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge Univ. Press, Cambridge, UK, 2 edition, 2004.
[9] Yang Liu, Xiao Yuan, Ming-Han Li, Weijun Zhang, Qi Zhao, Jiaqiang Zhong, Yuan Cao, Yu-Huai Li, Luo-Kan Chen, Hao Li, Tianyi Peng, Yu-Ao Chen, Cheng-Zhi Peng, Sheng-Cai Shi, Zhen Wang, Lixing You, Xiongfeng Ma, Jingyun Fan, Qiang Zhang, and Jian-Wei Pan. High-speed device-independent quantum random number generation without a detection loophole. *Phys. Rev. Lett.*, 120:010503, 2018.
[10] Lijiong Shen, Jianwei Lee, Le Phuc Tinh, Jean-Daniel Bancal, Alessandro Cer, Antia Lamas-Linares, Adriana Lita, Thomas Gerrits, Sae Woo Nam, Valerio Scarani, and Christian Kurtsiefer. Randomness extraction from Bell violation with continuous parametric down conversion. *Phys. Rev. Lett.*, 121:150402, Oct 2018.
[11] Yang Liu, Qi Zhao, Ming-Han Li, Jian-Yu Guan, Yanbao Zhang, Bing Bai, Weijun Zhang, Wen-Zhao Liu, Cheng Wu, Xiao Yuan, Hao Li, W. J. Munro, Zhen Wang, Lixing You, Jun Zhang, Xiongfeng Ma, Jingyun Fan, Qiang Zhang, and Jian-Wei Pan. Device independent quantum random number generation. *Nature*, 562:548–551, 2018.
[12] See Supplemental Material at http://link.aps.org/supplemental/x.xx/physrevlettxx.xxxx, which includes Refs. [25–38] for details.
[13] Yanbao Zhang, Honghao Fu, and Emanuel Knill. Efficient randomness certication by quantum probability estimation. submitted to Phys. Rev. Research (see arXiv:1806.04553 for an extended version), 2018.
[14] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. 23:880–884, 1969.
[15] L. Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–79, 2001.
[16] W. Mauerer, C. Portmann, and V. B. Scholz. A modular framework for randomness extraction based on Trevisan's construction. arXiv:1212.0520, code available on Github., 2012.
[17] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH, Zürich, Switzerland, 2006. quant-ph/0512258.
[18] M. Tomamichel. *Quantum Information Processing with Finite Resources - Mathematical Foundations*. SpringerBriefs in Mathematical Physics. Springer Verlag, 2016.
[19] C. A. Miller and Y. Shi. Universal security for randomness expansion from the spot-checking protocol. *SIAM J. Comput*, 46:1304–1335, 2017.

[20] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nature Communications*, 9:459, 2018.

[21] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellan, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam. A strong loophole-free test of local realism. 115:250402, 2015.

[22] Carlos Abellán, Waldimar Amaya, Daniel Mitrani, Valerio Pruneri, and Morgan W. Mitchell. Generation of fresh and pure random numbers for loophole-free Bell tests. *Phys. Rev. Lett.*, 115:250403, 2015.

[23] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam. Detecting single infrared photons with 93% system efficiency. *Nat. Photonics*, 7:210, 2013.

[24] J. Barrett, R. Colbeck, and A. Kent. Memory attacks on device-independent quantum cryptography. *Phys. Rev. Lett.*, 110(1):010503, 2013.

[25] Emanuel Knill, Yanbao Zhang, and Peter Bierhorst. Quantum randomness generation by probability estimation with classical side information. arXiv:1709.06159, 2017.

[26] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24(3):379–85, 1994.

[27] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, 2005.

[28] R. L. Frank and E. H. Lieb. Monotonicity of a relative Rényi entropy. *J. Math. Phys.*, 54:122201, 2013.

[29] M. Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory.* PhD thesis, ETH, Zürich, Switzerland, 2012. arXiv:1203.2142 (specific citations are for version 2).

[30] Anindya De, Christopher Portmann, Thomas Vidick, and Renato Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41(4):915–940, 2012.

[31] Xiongfeng Ma, Zhen Zhang, and Xiaoqing Tan. Explicit combinatorial design. arXiv:1109.6147, 2012.

[32] B. S. Cirelśon. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4:93, 1980.

[33] S. N. Bernstein. *Theory of Probability.* Moscow, 1927.

[34] W. van Dam, R. D. Gill, and P. D. Grunwald. The statistical strength of nonlocality proofs. *IEEE Trans. Inf. Theory.*, 51:2812–2835, 2005.

[35] Y. Zhang, E. Knill, and S. Glancy. Statistical strength of experiments to reject local realism with photon pairs and inefficient detectors. 81:032117/1–7, 2010.

[36] Yanbao Zhang, Emanuel Knill, and Peter Bierhorst. Certifying quantum randomness by probability estimation. *Phys. Rev. A*, 98:040304(R), Oct 2018.

[37] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, Jun 2007.

[38] Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New J. Phys.*, 11:045021, 2009.