Quantum Randomness from Probability Estimation with Classical Side Information

Yanbao Zhang,[*] Emanuel Knill,[†] Peter Bierhorst,[‡] and Scott Glancy[‡]

Random numbers can be generated from measurement results on suitably configured quantum devices. With care, the generated randomness is private. Given a configuration that demonstrates a violation of local realism, one can certify the generated randomness in a device-independent way [1]. However, even in the device-independent scenario, we still constrain the behavior of the randomness-generation device and the interaction between the device and any external entity by laws of physics. We also assume that initial randomness (which can be from a public source) is available in order to demonstrate the violation of local realism.

Depending on the specific models and assumptions, different methods for certifying device-independent randomness have been developed, see Refs. [2–7]. Almost all the previous methods certify randomness based only on the observed violation of one specific Bell inequality. We note that the two recent works [5] and [7] are exceptions. But their performance in the current experimental regime for photonic experiments with low violations of the CHSH Bell inequality per trial is unclear.

Recently in Ref. [8] we demonstrated that randomness can be certified from the photonic loophole-free Bell test reported in Ref. [9]. Motivated by this demonstration, here we propose a more powerful method to certify randomness.

Our contribution. Our method is based on what we call "probability estimation" (PE). The goal of PE is to obtain high-confidence-level upper bounds on the actual probability of observing the sequence of experimental results given known constraints on the distributions. We take advantage of the theory of test supermartingales to bypass the framework of Bell inequalities. From the full record of experimental results, we perform PE by computing products of "probability estimation factors" (PEFs). These factors provide a way of multiplicatively accumulating probability estimates trial-by-trial, where we view a Bell-test experiment as consisting of a sequence of trials, each with a (joint) setting input and a (joint) measurement output. We develop methods to construct PEFs when the set of distributions considered is convex. In particular, when the convex set is a polytope, optimal PEFs can be efficiently constructed using well-established optimization tools.

Our method works without assuming independent and identically distributed (i.i.d.) results from different trials in an experiment. It also allows for adaptive constructions of PEFs that can track changes during an experiment. This is helpful for the current experiments, where measurable drifts in state and setting parameters can wipe out a randomness certificate. Due to this adaptive feature, the number of trials need not be predetermined, and one can stop running the protocol as soon as the desired amount of randomness is extractable. Note that adaptiveness has also been exploited in our work on the analysis of tests of local realism [10, 11]. When the trial results are i.i.d., the bounds on the probability estimated and the corresponding bounds on the amount of randomness certified are asymptotically optimal given a constant error.

We apply our method to device-independent randomness-generation protocols. In particular, we consider protocols in which external entities have only classical side information. To certify randomness that is private with respect to such entities, we assume that the trial results satisfy non-signaling constraints. We can also add Tsirelson's bound to enforce "quantum constraints". Below we demonstrate our method with two examples, considering both sets of constraints.

First, we re-analyze the experimental results presented in Ref. [9] and analyzed in Ref. [8] where 256 random bits within 0.001 (in terms of the total-variation distance) of uniform were extracted with respect to classical side information and assuming non-signaling constraints. With PE, we can certify the presence of approximately twice as much randomness in the raw data with the same error parameter, or approximately four times as much if we assume the stronger quantum constraints, as shown in Fig. 1. In Ref. [8] we assumed that the setting distribution is uniform. In this experiment the setting choices slightly deviate from uniform, as discussed in Ref. [9]. We define the bias b of a random bit as twice the largest deviation from uniform in terms of the total-variation distance. The amounts of randomness certified at a few representative biases are shown in Fig. 1. We also re-analyzed the results presented in Ref. [2] where 42 random bits can be generated at the 99 % confidence level. With our method, it turns out that at least nine times more random bits can be generated at the same confidence

[*] NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan

[†] National Institute of Standards and Technology, Boulder, Colorado 80305, USA; Center for Theory of Quantum Matter, University of Colorado, Boulder, Colorado 80309, USA

[‡] National Institute of Standards and Technology, Boulder, Colorado 80305, USA

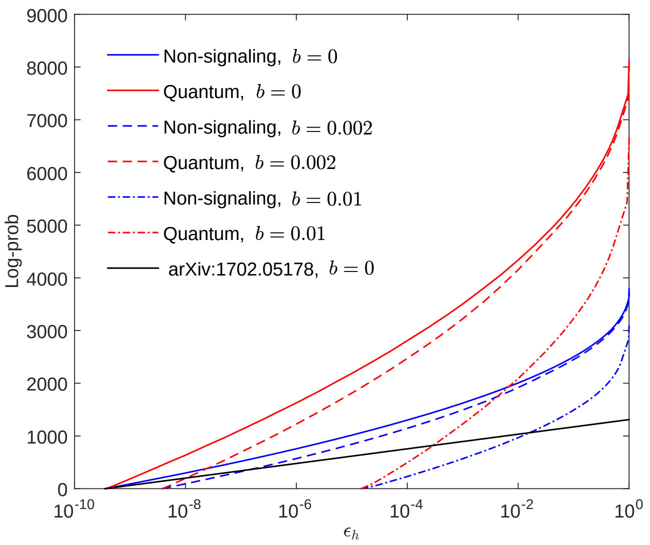

FIG. 1: Log-prob achieved (the amount of randomness certified) in the XOR 3 data set from Ref. [9]. The log-prob is the absolute value of logarithm with respect to base 2 of the final probability estimate achieved. The error bound ϵ_h is an input parameter. The curves show the achieved log-probs for non-signaling and quantum constraints at three representative biases. The curve that is lowest on the right is the log-prob reported in Ref. [8].

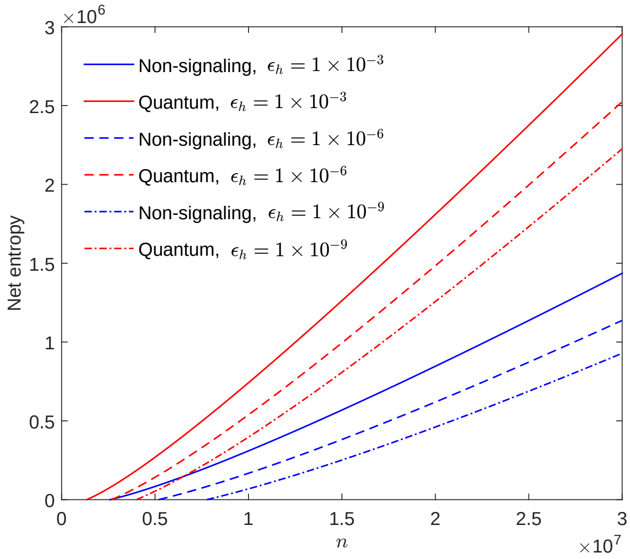

FIG. 2: Expected net entropy at ρ_{atoms}. We optimized the probability r of test trials given the number of trials n and the error bound ϵ_h.

level.

Second, we consider the challenge of producing more random bits than are consumed. i.e., randomness expansion. This requires a strategy to minimize the entropy used for the input setting choices. In each trial, with probability r we use a uniform setting distribution (a test trial) and with probability $(1-r)$, we use a fixed setting. Setting entropy is low if r is small. We use the outcome table S11 from the supplementary material of Ref. [12] to determine a representative distribution ρ_{atoms} for state-of-the-art loophole-free Bell tests involving entangled atoms. We assume that the distribution of measurement outcomes conditional on setting choices at each trial is given by ρ_{atoms}. We determine the net entropy according to PE after n trials in the protocol. Nominally, this is the net number of random bits that can be generated (without accounting for extractor parameters). The results are shown in Fig. 2.

Main idea of our method. We consider a randomness-generation protocol with input a random sequence $\mathbf{Z} = (Z_i)_{i=1}^N$ and output $\mathbf{C} = (C_i)_{i=1}^N$. Here, Z_i and C_i are the input and output of the device at the i'th trial. We assume that the number of possible values of each Z_i and C_i is finite for all i. We consider the situation where there is an external entity E who may have prepared (or accessed) the randomness-generation device before the user runs the protocol. Once the protocol starts, E has no further interaction with the device and holds only classical side information. The lack of interaction implies that any changes to E after running the protocol are purely external and independent of events inside the protocol given E's initial state. We can therefore time-shift these changes to the beginning of the protocol and thus remove any dynamics of E. This justifies the use of a single random variable E to describe the state of the external entity E.

Intuitively, the output of the device is random if the probability $\mathbb{P}(\mathbf{C}|\mathbf{Z}, E)$ is bounded away from 1 given constraints on the distributions at each trial. The distribution at a trial is conditional on E and the past information. We assume that all the possible distributions of $C_i Z_i$ at a trial, conditional on E and the past information, form a convex set, denoted by $\mathcal{H}$. We first propose a method to obtain a uniform upper bound on the probability $\mathbb{P}(\mathbf{C}|\mathbf{Z}, E = e)$ for all distributions in $\mathcal{H}$ and all possible values e of E. This is what we call "probability estimation". Later we discuss how to turn this upper bound into a formal statement on the amount of randomness certified.

Our method relies on a random sequence called a "test supermartingale" [13]. A test supermartingale with respect to a random sequence $\mathbf{R} = (R_i)_{i=0}^N$ is a random sequence $\mathbf{T} = (T_i)_{i=0}^N$ with the following properties: 1) $T_0 = 1$; 2) for all $i > 0$, $T_i \geq 0$; 3) T_i is determined by $\mathbf{R}_{\leq i}$; and 4) $\mathbb{E}(T_{i+1}|\mathbf{R}_{\leq i}) \leq T_i$ with respect to all distributions in $\mathcal{H}$. Here we denote $(R_i)_{i=0}^k$ by $\mathbf{R}_{\leq k}$. For the purpose of randomness generation, the sequence $\mathbf{R}$ captures all the relevant information that becomes available in the sequence of trials. In particular, R_i includes all the data that becomes available during the i'th trial. We refer to the ratio $F_i = T_i/T_{i-1}$ as a test factor. We can also define test supermartingales in terms of test factors: Let $\mathbf{F}$ be a random sequence such that for all i, $F_i \geq 0$, F_i is determined by $\mathbf{R}_{\leq i}$, and $\mathbb{E}(F_{i+1}|\mathbf{R}_{\leq i}) \leq 1$ with respect to all distributions in $\mathcal{H}$. Then the random sequence $\mathbf{T} = (T_i)_{i=0}^{N}$, defined by $T_0 = 1$ and $T_i = \prod_{j=1}^{i} F_j$ when $i > 0$, is a test supermartingale: Properties 1-3 above are immediate and by definition we can write $\mathbb{E}(T_{i+1}|\mathbf{R}_{\leq i}) = T_i\mathbb{E}(F_{i+1}|\mathbf{R}_{\leq i}) \leq T_i$. In this work, this is how we construct test supermartingales.

The definition implies that given a test supermartingale $\mathbf{T}$, $\mathbb{E}(T_i) \leq 1$ for all i. (This follows inductively from $\mathbb{E}(T_{i+1}) = \mathbb{E}(\mathbb{E}(T_{i+1}|\mathbf{R}_{\leq i})) \leq \mathbb{E}(T_i)$ and $T_0 = 1$.) An application of Markov's inequality then shows that for all distributions in $\mathcal{H}$ and $\epsilon_h > 0$,

$$\mathbb{P}(T_N \geq 1/\epsilon_h) \leq \epsilon_h. \tag{1}$$

Thus, a large final value $t = T_N$ of the test supermartingale is evidence against $\mathcal{H}$ with the p-value bound $1/t$ in the hypothesis test of the model $\mathcal{H}$ [10, 13].

For the purpose of PE, we construct a random sequence $\mathbf{T} = (T_i)_{i=1}^{N}$, where $T_i = \prod_{j=1}^{i} F_j$ and F_j is a deterministic function of the output C_j and input Z_j of the j'th trial, such that the product of the two random sequences $\mathbf{T}$ and $(\mathbb{P}(\mathbf{C}|\mathbf{Z}, E = e))^{\beta}$ with $\beta > 0$ is a test supermartingale for each e. That is, we have $\mathbb{E}(T_{i+1}(\mathbb{P}(\mathbf{C}_{\leq i+1}|\mathbf{Z}_{\leq i+1}, E = e))^{\beta}|\mathbf{R}_{\leq i}) \leq T_i(\mathbb{P}(\mathbf{C}_{\leq i}|\mathbf{Z}_{\leq i}, E = e))^{\beta}$ for all the distributions in $\mathcal{H}$. Since we do not know the probability $\mathbb{P}(\mathbf{C}_{\leq i}|\mathbf{Z}_{\leq i}, E = e)$ up to the i'th trial, we cannot determine the value of the test supermartingale at a trial given all the past information. We call this an "implicit test supermartingale". But, we can still apply Markov's inequality to show that for all distributions in $\mathcal{H}$ and $\epsilon_h > 0$,

$$\mathbb{P}(\mathbb{P}(\mathbf{C}|\mathbf{Z}, E = e) \geq (\epsilon_h T_N)^{-1/\beta}) \leq \epsilon_h. \tag{2}$$

The above inequality implies that the interval $(0, (\epsilon_h T_N)^{-1/\beta}]$ is a confidence interval for the probability $\mathbb{P}(\mathbf{C}|\mathbf{Z}, E = e)$ at the $(1-\epsilon_h)$ confidence level. In this sense, we call the factors F_i "probability estimation factors" (PEFs). We note that the bounds in Eqs. (1) and (2) can be further improved by using Doob's maximal inequality [14] instead of Markov's inequality.

Next we show how to construct PEFs. By the definition of test supermartingales, for all i the factors F_i need to satisfy the following two properties: 1) $F_i \geq 0$, and 2) $\mathbb{E}(F_i(\mathbb{P}(C_i|\mathbf{Z}_{\leq i}, \mathbf{C}_{\leq(i-1)}, E = e))^{\beta}) \leq 1$ with respect to all distributions in $\mathcal{H}$. When the set $\mathcal{H}$ is convex, we can prove that the second property is satisfied once it is satisfied with respect to all the extreme points of the set $\mathcal{H}$. Hence, when the set $\mathcal{H}$ has a finite number of extreme points, the second property implies a finite number of linear constraints on the factors F_i. To apply PE, we also require that the input Z_i at the i'th trial is independent of the past outputs $\mathbf{C}_{\leq(i-1)}$ given the classical side information E and the past inputs $\mathbf{Z}_{\leq(i-1)}$. This, along with the chain rule for conditional probabilities, allows us to demonstrate the equivalence of $\prod_{j=1}^{i} F_j(\mathbb{P}(C_j|\mathbf{Z}_{\leq j}, \mathbf{C}_{\leq(j-1)}, E = e))^{\beta}$ and $T_i(\mathbb{P}(\mathbf{C}_{\leq i}|\mathbf{Z}_{\leq i}, E = e))^{\beta}$.

To minimize the typical upper bound on the probability $\mathbb{P}(\mathbf{C}|\mathbf{Z}, E = e)$, before observing the results Z_i and C_i at the i'th trial we can construct F_i so as to maximize the expectation $\mathbb{E}_{\mu}\log_2(F_i)$ given the estimated distribution μ before the i'th trial. This optimization promises asymptotic optimality in the i.i.d. case for a constant error bound. Note that we also need to determine the value of β before running the protocol.

To determine the amount of extractable randomness, rather than determining a smoothed min-entropy, we directly compose the probability estimate with a suitable extractor. For this, we assume that the protocol is configured to produce a requested number σ of random bits within ϵ_h of uniform. If banked randomness is available, then we can take a probability estimate $2^{-\sigma_h}$ with error bound ϵ_h (i.e., with the $(1-\epsilon_h)$ confidence level) for the output sequence. If $\sigma_h < \sigma$, we fill in the output with $\lceil \sigma - \sigma_h \rceil$ bits and feed this into an extractor. In this case, the filled output has ϵ_h-smoothed min-entropy σ, so the composition is straightforward. In the absence of banked randomness, the protocol has a probability of failure, and a composition and analysis similar to that in Ref. [8] results in the expected number of random bits (up to adjustments for extractor parameters) when the protocol succeeds. Either way, the probability estimate translates directly into the number of extracted bits.

We remark that our framework is in the spirit of the entropy-accumulation framework of Ref. [5, 6], but takes advantage of the simplifications possible for randomness generation with respect to classical side information. In the entropy-accumulation framework, the relevant estimators, called min-tradeoff functions, must be chosen before the protocol, and the final certificate is based on a sum of statistics derived from these functions. The error bound affects the certificate non-linearly and has a significant impact on the amount of randomness certified. In the PE framework, PEFs can be adapted, and probability estimators accumulate multiplicatively. For relevant situations, PEFs can be readily optimized. The error bound reduces the probability estimate by a straightforward multiplicative factor, and hence the amount of randomness certified is reduced additively, yielding a better controlled error bound vs. randomness tradeoff.

[1] R. Colbeck, Ph.D. thesis, University of Cambridge (2007).

[2] 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, et al., Nature **464**, 1021 (2010).

[3] U. Vazirani and T. Vidick, in *STOC'12 Proceedings of the 44th Annual ACM Symposium on Theory of Computing* (2012), p. 61.

[4] C. Miller and Y. Shi, in *STOC'14 Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (2014), pp. 417–426.

[5] F. Dupuis, O. Fawzi, and R. Renner (2016), arXiv:1607.01796.

[6] R. Arnon-Friedman, R. Renner, and T. Vidick (2016), arXiv:1607.01797.

[7] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio (2016), arXiv:1611.00352.

[8] P. Bierhorst, E. Knill, S. Glancy, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, and L. K. Shalm (2017), arXiv:1702.05178.

[9] 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, et al., Phys. Rev. Lett. **115**, 250402 (2015).

[10] Y. Zhang, S. Glancy, and E. Knill, Phys. Rev. A **84**, 062118 (2011).

[11] Y. Zhang, S. Glancy, and E. Knill, Phys. Rev. A **88**, 052119 (2013).

[12] W. Rosenfeld, D. Burchardt, R. Garthoff, K. Redeker, N. Ortegel, and M. R. ahd H. Weinfurter (2016), arXiv:1611.04604.

[13] G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk, Statistical Science **26**, 84 (2011).

[14] C. Dellacherie and P. Meyer, *Probabilities and Potential, B: Theory of Martingales*, North-Holland Mathematics Studies (Elsevier Science, 2011), ISBN 9780080871837.