Certifying quantum randomness by probability estimation

Yanbao Zhang, Emanuel Knill, and Peter Bierhorst

Phys. Rev. A **98**, 040304 — Published 31 October 2018

DOI: 10.1103/PhysRevA.98.040304

Certifying Quantum Randomness by Probability Estimation

Yanbao Zhang,[1,*] Emanuel Knill,[2,3] and Peter Bierhorst[2]

[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] *Center for Theory of Quantum Matter, University of Colorado, Boulder, Colorado 80309, USA*

We introduce probability estimation, a broadly applicable framework to certify randomness in a finite sequence of measurements subject to verifiable physical constraints and with respect to classical side information. Examples include randomness from single-photon measurements and device-independent randomness from Bell tests. Advantages of probability estimation include unproblematic early stopping when goals are achieved, optimal randomness rates, applicability to Bell tests with small violations, and unsurpassed finite-data efficiency. We greatly reduce latencies for producing random bits and formulate an associated rate-tradeoff problem of independent interest. We also show that the latency is determined by an information-theoretic measure of nonlocality rather than the Bell violation.

Randomness is a key enabling resource for computation and communication. Besides being required for Monte-Carlo simulations and statistical sampling, private random bits are needed for initiating authenticated connections and establishing shared keys, both common tasks for browsers, servers and other online entities [1]. Public random bits from "randomness beacons" have applications to fair resource sharing [2] and can seed private randomness sources based on quantum mechanics [3]. Common requirements for random bits are that they are unpredictable to all before they are generated, and private to the users before they are published.

Quantum mechanics provides natural opportunities for generating randomness. The best known example involves measuring a two-level system that is in an equal superposition of its two levels. A disadvantage of such schemes is that they require trust in the measurement apparatus, and undiagnosed failures are always a possibility. This disadvantage is overcome by a loophole-free Bell test [4, 5], which can generate output whose randomness can be certified solely by statistical tests of setting and outcome frequencies. The devices preparing the quantum states and performing the measurements may come from an untrusted source. This strategy for certified randomness generation is known as device-independent randomness generation (DIRG).

Loophole-free Bell tests have been realized with nitrogen-vacancy (NV) centers [6], with atoms [7] and with photons [8, 9], enabling the possibility of full experimental implementations of DIRG. However, for NV centers and atoms, the rate of trials is too low, and for photons, the violation per trial is too small. As a result, previously available DIRG protocols [3, 10–18] are not ready for implementation with current loophole-free Bell tests. These protocols do not achieve good finite-data efficiency and therefore require an impractical number of trials. Experimental techniques will improve, but for many applications of randomness generation, including randomness beacons and key generation, it is desirable to achieve finite-data efficiency that is as high as possible, since these applications often require short blocks of fresh random bits with minimum delay or latency.

Excellent finite-data efficiency was achieved by a method that we described and implemented in Refs. [19, 20], which reduced the time required for generating 1024 low-error random bits with respect to classical side information from hours to minutes for a state-of-the-art photonic loophole-free Bell test. The method in Refs. [19, 20] is based on the prediction-based ratio (PBR) analysis [21] for hypothesis tests of local realism. Specifically, in Refs. [19, 20] we established a connection between the PBR-based p-value and the amount of randomness certified against classical side information. The basis for success of the method of Refs. [19, 20] motivates our development of probability estimation for randomness certification, with better finite-data efficiency and with broader applications.

In the probability estimation framework, the amount of certified randomness is *directly* estimated without relying on hypothesis tests of local realism. To certify randomness, we first obtain a bound on the conditional probability of the observed outcomes given the chosen settings, valid for all classical side information. Then we show how to obtain conditional entropy estimates from this bound to quantify the number of extractable random bits [22]. By focusing on data-dependent probability estimates, we are able to take advantage of powerful statistical techniques to obtain the desired bound. The statistical techniques are based on test supermartingales [23] and Markov's bounds. Probability estimation inherits several features of the theory of test supermartingales. For example, probability estimation has no independence or stationarity requirement on the probability distribution of trial results. Also, probability estimation supports stopping the experiment early, as soon as the randomness goal is achieved.

Probability estimation is broadly applicable. In particular it is not limited to device-independent scenarios and

* Corresponding author.
yanbao.zhang@lab.ntt.co.jp

can be applied to traditional randomness generation with quantum devices. Such applications are enabled by the notion of models, which are sets of probability distributions that capture verified, physical constraints on device behavior. In the case of Bell tests, these constraints include the familiar non-signaling conditions [24, 25]. In the case of two-level systems such as polarized photons, the constraints can capture that measurement angles are within a known range, for example.

In this rapid communication, we first describe the technical features of probability estimation and the main results that enable its practical use. We propose a general information-theoretic rate-tradeoff problem that closely relates to finite-data efficiency. We then show how the general theoretical concepts are instantiated in experimentally relevant examples involving Bell-test configurations. We demonstrate advantages of probability estimation such as its optimal asymptotic randomness rates and show large improvements in finite-data efficiency, which corresponds to great reductions in latency.

Theory. Consider an experiment with "inputs" $\mathbf{Z}$ and "outputs" $\mathbf{C}$. The inputs normally consist of the random choices made for measurement settings but may include choices of state preparations such as in the protocols of Refs. [26, 27]. The outputs consist of the corresponding measurement outcomes. In the cases of interest, the inputs and outputs are obtained in a sequence of n time-ordered trials, where the i'th trial has input Z_i and output C_i, and $\mathbf{Z} = (Z_i)_{i=1}^n$ and $\mathbf{C} = (C_i)_{i=1}^n$. We assume that Z_i and C_i are countable-valued. We refer to the trial inputs and outputs collectively as the trial "results", and to the trials preceding the upcoming one as the "past". The party with respect to which the randomness is intended to be unpredictable is represented by an external classical system, whose initial state before the experiment may be correlated with the devices used. The classical system carries the side information E, which is assumed to be countable-valued. After the experiment, the joint of $\mathbf{Z}$, $\mathbf{C}$ and E is described by a probability distribution μ. The upper-case symbols introduced in this paragraph are treated as random variables. As is conventional, their values are denoted by the corresponding lower-case symbols.

The amount of extractable uniform randomness in $\mathbf{C}$ conditional on both $\mathbf{Z}$ and E is quantified by the (classical) smooth conditional min-entropy $H_{\min}^{\epsilon}(\mathbf{C}|\mathbf{Z}E)_\mu$ where ϵ is the "error bound" (or "smoothness") and μ is the joint distribution of $\mathbf{Z}$, $\mathbf{C}$ and E. One way to define the smooth conditional min-entropy is with the conditional guessing probability $P_{\text{guess}}(\mathbf{C}|\mathbf{Z}E)_\mu$ defined as the average over values $\mathbf{z}$ and e of the maximum conditional probability $\max_{\mathbf{c}} \mu(\mathbf{c}|\mathbf{z}e)$. The ϵ-smooth conditional min-entropy $H_{\min}^{\epsilon}(\mathbf{C}|\mathbf{Z}E)_\mu$ is the greatest lower bound of $-\log_2 P_{\text{guess}}(\mathbf{C}|\mathbf{Z}E)_{\mu'}$ for all distributions μ' within total-variation distance ϵ of μ. Our goal is to obtain lower bounds on $H_{\min}^{\epsilon}(\mathbf{C}|\mathbf{Z}E)_\mu$ with probability estimation.

The application of probability estimation requires a notion of models. A model $\mathcal{H}$ for an experiment is defined as the set of all probability distributions of $\mathbf{Z}$ and $\mathbf{C}$ achievable in the experiment conditionally on values e of E. If a joint distribution μ of $\mathbf{Z}$, $\mathbf{C}$ and E satisfies that for all e, the conditional distributions $\mu(\mathbf{C}\mathbf{Z}|e)$, considered as distributions of $\mathbf{Z}$ and $\mathbf{C}$, are in $\mathcal{H}$, we say that the distribution μ satisfies the model $\mathcal{H}$.

To apply probability estimation to an experiment consisting of n time-ordered trials, we construct the model $\mathcal{H}$ for the experiment as a chain of models $\mathcal{C}_i$ for each individual trial i in the experiment. The trial model $\mathcal{C}_i$ is defined as the set of all probability distributions of trial results C_iZ_i achievable at the i'th trial conditionally on both the past trial results and the side information E. For example, for Bell tests, $\mathcal{C}_i$ may be the set of non-signaling distributions with uniformly random inputs. Let $\mathbf{z}_{<i} = (z_j)_{j=1}^{i-1}$ and $\mathbf{c}_{<i} = (c_j)_{j=1}^{i-1}$ be the results before the i'th trial. The sequences $\mathbf{z}_{\leq i}$ and $\mathbf{c}_{\leq i}$ are defined similarly. The chained model $\mathcal{H}$ consists of all conditional distributions $\mu(\mathbf{C}\mathbf{Z}|e)$ satisfying the following two conditions. First, at each trial i the conditional distributions $\mu(C_iZ_i|\mathbf{c}_{<i}\mathbf{z}_{<i}e)$ for all $\mathbf{c}_{<i}$, $\mathbf{z}_{<i}$ and e are in the trial model $\mathcal{C}_i$. Second, at each trial i the input Z_i is independent of the past outputs $\mathbf{C}_{<i}$ given E and the past inputs $\mathbf{Z}_{<i}$. The second condition prevents leaking information about the past outputs through the future inputs, which is necessary for certifying randomness in the outputs $\mathbf{C}$ conditional on both the inputs $\mathbf{Z}$ and the side information E. In the common situation where the inputs are chosen independently with distributions known before the experiment, the second condition is always satisfied.

Since the model $\mathcal{H}$ consists of all conditional distributions $\mu(\mathbf{C}\mathbf{Z}|e)$ regardless of the value e, the analyses in the next paragraph apply to the worst-case conditional distribution over e. To simplify notation we normally write the distribution $\mu(\mathbf{C}\mathbf{Z}|e)$ conditional on e as $\mu_e(\mathbf{C}\mathbf{Z})$, abbreviated as μ_e.

To estimate the conditional probability $\mu_e(\mathbf{c}|\mathbf{z})$, we design trial-wise probability estimation factors (PEFs) and multiply them. Consider a generic trial with trial model $\mathcal{C}$, where for generic trials, we omit the trial index. Let $\beta > 0$. A PEF with power β for $\mathcal{C}$ is a function $F : cz \mapsto F(cz) \geq 0$ such that for all $\sigma \in \mathcal{C}$, $\mathbb{E}_\sigma\big(F(CZ)\sigma(C|Z)^\beta\big) \leq 1$, where $\mathbb{E}$ denotes the expectation functional. Note that $F(cz) = 1$ for all cz defines a valid PEF with each positive power. For each i, let F_i be a PEF with power β for the i'th trial, where the PEF can be chosen adaptively based on the past results $\mathbf{c}_{<i}\mathbf{z}_{<i}$. Other information from the past may also be used, see Ref. [28]. Let $T_0 = 1$ and $T_i = \prod_{j=1}^{i} F_j(C_jZ_j)$. The final value T_n of the running product T_i, where n is the total number of trials in the experiment, determines the probability estimate. Specifically, for each value e of E, each μ_e in the chained model $\mathcal{H}$, and $\epsilon > 0$, we have

$$\mathbb{P}_{\mu_e}\big(\mu_e(\mathbf{C}|\mathbf{Z}) \geq U(\mathbf{C}\mathbf{Z})\big) \leq \epsilon, \tag{1}$$

where $\mathbb{P}_{\mu_e}$ denotes the probability according to the distribution μ_e and $U(\mathbf{CZ}) = (\epsilon T_n)^{-1/\beta}$. The proof of Eq. (1) is given in the Supplemental Material (SM) [29]. The meaning of Eq. (1) is as follows: For each e and each $\mu_e \in \mathcal{H}$, the probability that $\mathbf{C}$ and $\mathbf{Z}$ take values $\mathbf{c}$ and $\mathbf{z}$ for which $U(\mathbf{C} = \mathbf{c}, \mathbf{Z} = \mathbf{z}) \leq \mu_e(\mathbf{C} = \mathbf{c}|\mathbf{Z} = \mathbf{z})$ is at most ϵ. This defines $U(\mathbf{CZ}) = (\epsilon T_n)^{-1/\beta}$ as a level-ϵ probability estimator.

A main theorem of probability estimation is the connection between probability estimators and conditional min-entropy estimators, which is formalized as follows:

Theorem 1. *Suppose that the joint distribution μ of $\mathbf{Z}$, $\mathbf{C}$ and E satisfies the chained model $\mathcal{H}$. Let $1 \geq \kappa, \epsilon > 0$ and $1 \geq p \geq 1/|\mathrm{Rng}(\mathbf{C})|$, where $|\mathrm{Rng}(\mathbf{C})|$ is the number of possible outputs. Define $\{\phi\}$ to be the event that $T_n \geq 1/(p^\beta \epsilon)$, and let $\kappa \leq \mathbb{P}_\mu(\phi)$. Then the smooth conditional min-entropy satisfies*

$$H_{\min}^{\epsilon}(\mathbf{C}|\mathbf{Z}E;\phi) \geq -\log_2(p/\kappa^{1+1/\beta}).$$

The probability of the event $\{\phi\}$ can be interpreted as the probability that the experiment succeeds, and κ is an assumed lower bound on the success probability. The theorem is proven in SM [29].

When constructing PEFs, the power $\beta > 0$ must be decided *before* the experiment and cannot be adapted. Thm. 1 requires that p, ϵ and κ also be chosen *beforehand*, and success of the experiment requires $T_n \geq 1/(p^\beta \epsilon)$, or equivalently,

$$\log_2(T_n)/\beta + \log_2(\epsilon)/\beta \geq -\log_2(p). \tag{2}$$

Since $\log_2(T_n) = \sum_i \log_2(F_i)$, *before* the experiment we choose PEFs in order to aim for large expected values of the logarithms of the PEFs F_i. Consider a generic next trial with results CZ and model $\mathcal{C}$. Based on prior calibrations or the frequencies of observed results in past trials, we can determine a distribution $\nu \in \mathcal{C}$ that is a good approximation to the distribution of the next trial's results CZ. Many experiments are designed so that each trial's distribution is close to ν. The PEF can be optimized for this distribution but, by definition, is valid regardless of the actual distribution of the next trial in $\mathcal{C}$. Thus, one way to optimize PEFs *before* the next trial is as follows:

$$\begin{aligned} &\text{Max:} \quad \mathbb{E}_\nu\big(n\log_2(F(CZ))/\beta + \log_2(\epsilon)/\beta\big) \\ &\text{With:} \quad \textstyle\sum_{cz} F(cz)\sigma(c|z)^\beta\sigma(cz) \leq 1 \text{ for all } \sigma \in \mathcal{C}, \\ &\qquad F(cz) \geq 0, \text{ for all } cz. \end{aligned} \tag{3}$$

The objective function is strictly concave and the constraints are linear, so there is a unique maximum, which can be found by convex programming. More details are available in SM [29].

Before the experiment, one can also optimize the objective function in Eq. (3) with respect to the power β. During the experiment ϵ and β are fixed, so it suffices to maximize $\mathbb{E}_\nu\big(\log_2(F(CZ))\big)$. If during the experiment, the running product T_i with $i < n$ exceeds the target $1/(p^\beta\epsilon)$, we can set future PEFs to $F(CZ) = 1$, which is a valid PEF with power β. This ensures that $T_n = T_i$ and is equivalent to stopping the experiment after trial i. Since the target needs to be set conservatively in order to make the actual experiment succeed with high probability, this can result in a significant reduction in the number of trials actually executed.

A question is how PEFs perform asymptotically for a stable experiment. This question is answered by determining the rate per trial of entropy production assuming constant ϵ and κ independent of the number of trials. In view of Thm. 1, after n trials the entropy rate is given by $\big(-\log_2(p) + \log_2(\kappa^{1+1/\beta})\big)/n$. Considering Eq. (2), when n is large the entropy rate is dominated by $\log_2(T_n)/(n\beta)$, which is equal to $\sum_{i=1}^n \log_2(F_i)/(n\beta)$. Therefore, if each trial has distribution ν and each trial model is the same $\mathcal{C}$, then in the limit of large n the asymptotic entropy rate witnessed by a PEF F with power β is given by $\mathbb{E}_\nu\big(\log_2(F(CZ))/\beta\big)$. Define the rate

$$g(\beta) = \sup_F \mathbb{E}_\nu\big(\log_2(F(CZ))/\beta\big), \tag{4}$$

where the supremum is over PEFs F with power β for $\mathcal{C}$. The maximum asymptotic entropy rate at constant ϵ and κ witnessed by PEFs is $g_0 = \sup_{\beta>0} g(\beta)$. The rate $g(\beta)$ is non-increasing in β (SM [29]), so g_0 is determined by the limit as β goes to zero. A theorem proven in Ref. [28] is that g_0 is the worst-case conditional entropy $H(C|ZE)$ over joint distributions of CZE allowed by $\mathcal{C}$ with marginal ν. Since this is a tight upper bound on the asymptotic randomness rate [30], probability estimation is asymptotically optimal and we identify g_0 as the asymptotic randomness rate.

For finite data and applications requiring fresh blocks of randomness, the rate g_0 is not achieved. To understand why, consider the problem of certifying a fixed number of bits b of randomness at error bound ϵ and with as few trials as possible, where each trial has distribution ν. In view of Thm. 1, the PEF optimization problem in Eq. (3), and the definition of $g(\beta)$ in Eq. (4), n needs to be sufficiently large so that

$$ng(\beta) + \log_2(\epsilon)/\beta + (1 + 1/\beta)\log_2(\kappa) \geq b. \tag{5}$$

The left-hand side is maximized at positive β, whereas $g(\beta)$ increases to g_0 as β goes to zero. As a result the best actual rate b/n is less than g_0.

Setting $\kappa = 1$ in Eq. (5) shows that the number of trials n must exceed $-\log_2(\epsilon)/(\beta g(\beta))$ before randomness can be produced, which suggests that the maximum of $\beta g(\beta)$ is a good indicator of finite-data performance. Another way to arrive at this quantity is to consider $\epsilon = 2^{-\gamma n}$, where $\gamma > 0$ is the "certificate rate". Given ν and the trial model, we can ask for the maximum certificate rate for which it is possible to have positive entropy rate at $\kappa = 1$. It follows from Eq. (5) with $\kappa = 1$ that this rate

is at most

$$\gamma_{\mathrm{PEF}} = \sup_{\beta>0} \beta g(\beta). \tag{6}$$

We propose a general information-theoretic rate-tradeoff problem given trial model $\mathcal{C}$ and $\nu \in \mathcal{C}$: For a given certificate rate γ, determine the supremum of the entropy rates achievable by protocols. Eq. (5) implies lower bounds on the resulting tradeoff curve.

Our protocol assumes classical-only side information. There are more costly DIRG protocols that handle quantum side information [11, 13–17], but verifying that side information is effectively classical only requires confirming that the quantum devices used in the experiment have no long-term quantum memory. Verifying the absence of long-term quantum memory in current experiments is possibly less difficult than ensuring that there are no backdoors or information leaks in the experiment's hardware and software.

Applications. We consider DIRG with the standard two-party, two-setting, two-outcome Bell-test configuration [31]. The parties are labeled A and B. In each trial, a source prepares a state shared between the parties, and each party chooses a random setting (their input) and obtains a measurement outcome (their output). We write $Z = XY$, where X and Y are the inputs of A and B, and $C = AB$, where A and B are the respective outputs. For this configuration, $A, B, X, Y \in \{0, 1\}$.

Consider the trial model $\mathcal{N}$ consisting of distributions of $ABXY$ with uniformly random inputs and satisfying non-signaling [24]. We begin by determining and comparing the asymptotic randomness rates witnessed by different methods. The rates are usually quantified as functions of the expectation $\hat{I}$ of the CHSH Bell function (see Eq. (S27) in SM [29]) for $\hat{I} > 2$ (the classical upper bound). We prove in SM [29] that the maximum asymptotic randomness rate for any $\nu \in \mathcal{N}$ is equal to $(\hat{I}-2)/2$, and the rate g_0 witnessed by PEFs matches this value. Most previous studies, such as Refs. [3, 10, 12, 18, 32–34], estimate the asymptotic randomness rate by the single-trial conditional min-entropy $H_{\min}(AB|XYE)$. We determine that $H_{\min}(AB|XYE) = -\log_2((6-\hat{I})/4) < g_0$ when $2 < \hat{I} < 4$. As $\hat{I}$ decreases to 2 the ratio of g_0 to $H_{\min}(AB|XYE)$ approaches 1.386, demonstrating an improvement at small violations.

Next, we investigate finite-data performance. We consider three different families of quantum-achievable distributions of trial results. For the first family $\nu_{\mathrm{E},\theta}$, A and B share the unbalanced Bell state $|\Psi_\theta\rangle = \cos\theta|00\rangle + \sin\theta|11\rangle$ with $\theta \in (0, \pi/4]$ and apply projective measurements that maximize $\hat{I}$. This determines $\nu_{\mathrm{E},\theta}$. This family contains the goal states for many experiments suffering from detector inefficiency. For the second family $\nu_{\mathrm{W},p}$, A and B share a Werner state $\rho = p|\Psi_{\pi/4}\rangle\langle\Psi_{\pi/4}| + (1-p)\mathbb{1}/4$ with $p \in (1/\sqrt{2}, 1]$ and again apply measurements that maximize $\hat{I}$. Werner states are standard examples in quantum information and are among the worst states for our application. In experiments with photons, measurements are implemented with imperfect detectors. For the third family $\nu_{\mathrm{P},\eta}$, A and B use detectors with efficiency $\eta \in (2/3, 1)$ to implement the measurements and to close the detection loophole [35]. They choose the unbalanced Bell state $|\Psi_\theta\rangle$ and measurements such that an information-theoretic measure of nonlocality, the statistical strength for rejecting local realism [36–38], is maximized.

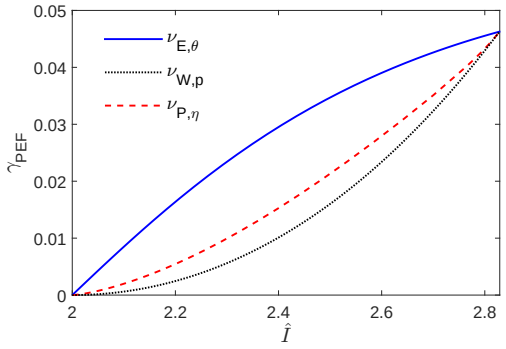

FIG. 1. Maximum certificate rates γ_{PEF} (Eq. (6)) as a function of $\hat{I}$ for each family of distributions.

For each family of distributions, we determine the maximum certificate rate γ_{PEF} as given in Eq. (6). For this, we consider the trial model $\mathcal{N}$, but we note that γ_{PEF} does not depend on the specific constraints on the quantum-achievable conditional distributions $\mathbb{P}(AB|XY)$ (SM [29]). As an indicator of finite-data performance, γ_{PEF} depends not only on $\hat{I}$, but also on the distribution ν. To illustrate this behavior, we plot the rates γ_{PEF} as a function of $\hat{I}$ for each family of distributions in Fig. 1. To obtain these plots, we note that $\hat{I}$ is a monotonic function of the parameter θ, p or η for each family. We also find that γ_{PEF} is given by the statistical strength of the distribution ν for rejecting local realism (see SM [29] for a proof). Conventionally, experiments are designed to maximize $\hat{I}$, but in general, the optimal state and measurements maximizing $\hat{I}$ are different from those maximizing the statistical strength [37, 38].

We further determine the minimum number of trials, $n_{\mathrm{PEF},b}$, required to certify b bits of ϵ-smooth conditional min-entropy with a given distribution ν of trial results. From Eq. (5), we get

$$n_{\mathrm{PEF},b} = \inf_{\beta>0} \frac{b\beta - \log_2(\epsilon) - (1+\beta)\log_2(\kappa)}{\beta g(\beta)},$$

where for simplicity we allow non-integer values for $n_{\mathrm{PEF},b}$. We can upper bound $n_{\mathrm{PEF},b}$ by means of the simpler-to-compute certificate rate γ_{PEF} given in Eq. (6). For the trial model $\mathcal{N}$, γ_{PEF} is achieved when β is above a threshold β_0 that depends on ν (SM [29]). From γ_{PEF} and β_0, we can determine the upper bound

$$n'_{\mathrm{PEF},b} = \big(b\beta_0 - \log_2(\epsilon) - (1+\beta_0)\log_2(\kappa)\big)/\gamma_{\mathrm{PEF}}$$

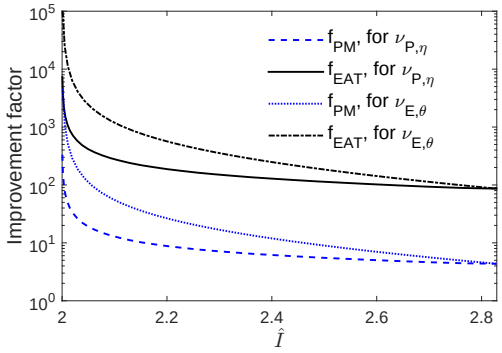

FIG. 2. Improvement factors as a function of $\hat{I}$.

on $n_{\mathrm{PEF},b}$. The minimum number of trials required can be determined for other published protocols, which usually certify conditional min-entropy from $\hat{I}$. (An exception is Ref. [18] but the minimum number of trials required is worse.) We consider the protocol "PM" of Ref. [3] and the entropy accumulation protocol "EAT" of Ref. [17]. From Thm. 1 of Ref. [3] with $\kappa = 1$ and $b \searrow 0$, we obtain a lower bound

$$n_{\mathrm{PM},0} = -2\log_e(\epsilon)/\big((\hat{I}-2)/(4+2\sqrt{2})\big)^2.$$

For the EAT protocol, we determine an explicit lower bound $n_{\mathrm{EAT},b}$ in SM [29]. This lower bound applies for $b \geq 0$ and $\epsilon, \kappa \in (0,1]$, and is valid with respect to quantum side information for the trial model consisting of quantum-achievable distributions.

We compare the three protocols over a broad range of $\hat{I}$ for $b \searrow 0$, $\epsilon = 10^{-6}$, and $\kappa = 1$. For each family of distributions above, we compute the improvement factors given by $f_{\mathrm{PM}} = n_{\mathrm{PM},0}/n'_{\mathrm{PEF},0}$ and $f_{\mathrm{EAT}} = n_{\mathrm{EAT},0}/n'_{\mathrm{PEF},0}$. For $\nu_{\mathrm{W},p}$, the improvement factors depend weakly on $\hat{I}$: f_{PM} increases from 3.89 at $\hat{I} = 2.008$ to 4.36 at $\hat{I} = 2\sqrt{2}$, while f_{EAT} increases from 84.97 at $\hat{I} = 2.008$ to 86.35 at $\hat{I} = 2\sqrt{2}$. For $\nu_{\mathrm{E},\theta}$ and $\nu_{\mathrm{P},\eta}$, the improvement factors can be much larger and depend strongly on $\hat{I}$, monotonically decreasing with $\hat{I}$ as shown in Fig. 2. The improvement is particularly notable at small violations which are typical in current photonic loophole-free Bell tests. We remark that similar comparison results were obtained with other choices of the values for ϵ and κ.

The large latency reduction with probability estimation persists for certifying blocks of randomness. For randomness beacons, good reference values are $b = 512$ and $\epsilon = 2^{-64}$. We also set $\kappa = 2^{-64}$. Setting $\kappa = \epsilon$ is a common conservative choice, but we remark that soundness for randomness generation can be defined with a better tradeoff between ϵ and κ [28]. We consider the trial model $\mathcal{T}$ of distributions with uniformly random inputs, satisfying both non-signaling conditions [24] and Tsirelson's bounds [39]. Consider the state-of-the-art photonic loophole-free Bell test reported in Ref. [20]. With probability estimation, the number of trials required for the distribution inferred from the measurement statistics is 4.668×10^7, which would require about 7.78 minutes of running time in the referenced experiment. With entropy accumulation [17], 2.887×10^{11} trials taking 802 hours would be required. We also reanalyzed the experimental data from Refs. [10] and [19] obtaining substantially better results with probability estimation, see SM [29] for details.

In conclusion, probability estimation is a powerful and flexible framework for certifying randomness in data from a finite sequence of experimental trials. Implemented with probability estimation factors, it witnesses optimal asymptotic randomness rates. For practical applications requiring fixed-size blocks of random bits, it can reduce the latencies by orders of magnitude even for high-quality devices. Latency is a notable problem for device-independent quantum key generation (DIQKD). If probability estimation can be extended to accommodate security against quantum side information, the latency reductions may be extendable to DIQKD by means of existing constructions [17].

ACKNOWLEDGMENTS

We thank D. N. Matsukevich for providing the experimental data for Ref. [10], Bill Munro, Carl Miller, Kevin Coakley, and Paulina Kuo for help with reviewing this paper. This work includes contributions of the National Institute of Standards and Technology, which are not subject to U.S. copyright.

[1] Christof Paar and Jan Pelzl, *Understanding Crypotgraphy* (Springer-Verlag Berlin Heidelberg, New York, 2010).
[2] M. J. Fischer, "A public randomness service," in *SECRYPT 2011* (2011) pp. 434–438.
[3] S. Pironio and S. Massar, "Security of practical private randomness generation," Phys. Rev. A **87**, 012336 (2013).
[4] R. Colbeck, *Quantum and Relativistic Protocols for Secure Multi-Party Computation*, Ph.D. thesis, University of Cambridge (2007).
[5] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," J. Phys. A: Math. Theor. **44**, 095305 (2011).
[6] B. Hensen *et al.*, "Loophole-free Bell inequality violation using electron spins separated by 1.3 km," Nature **526**, 682 (2015).
[7] W. Rosenfeld, D. Burchardt, R. Garthoff, K. Redeker, N. Ortegel, M. Rau, and H. Weinfurter, "Event-ready Bell-test using entangled atoms simultaneously closing detection and locality loopholes," Phys. Rev. Lett. **119**, 010402 (2017).

[8] M. Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger, "Significant-loophole-free test of Bell's theorem with entangled photons," Phys. Rev. Lett. **115**, 250401 (2015).

[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, 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. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong loophole-free test of local realism," Phys. Rev. Lett. **115**, 250402 (2015).

[10] 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," Nature **464**, 1021–4 (2010).

[11] U. Vazirani and T. Vidick, "Certifiable quantum dice - or, exponential randomness expansion," in *STOC'12 Proceedings of the 44th Annual ACM Symposium on Theory of Computing* (2012) p. 61.

[12] S. Fehr, R. Gelles, and C. Schaffner, "Security and composability of randomness expansion from Bell inequalities," Phys. Rev. A **87**, 012335 (2013).

[13] C. A. Miller and Y. Shi, "Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices," J. ACM **63**, 33 (2016).

[14] C. A. Miller and Y. Shi, "Universal security for randomness expansion from the spot-checking protocol," SIAM J. Comput **46**, 1304–1335 (2017).

[15] K.-M. Chung, Y. Shi, and X. Wu, "Physical randomness extractors: Generating random numbers with minimal assumptions," (2014), arXiv:1402.4797 [quant-ph].

[16] M. Coudron and H. Yuen, "Infinite randomness expansion with a constant number of devices," in *STOC'14 Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (2014) pp. 427–36.

[17] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, "Practical device-independent quantum cryptography via entropy accumulation," Nat. Commun. **9**, 459 (2018).

[18] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, "Device-independent randomness generation from several Bell estimators," New J. Phys. **20**, 023049 (2018).

[19] P. Bierhorst, E. Knill, S. Glancy, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, and L. K. Shalm, "Experimentally generated random numbers certified by the impossibility of superluminal signaling," (2017), arXiv:1702.05178.

[20] 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).

[21] Y. Zhang, S. Glancy, and E. Knill, "Asymptotically optimal data analysis for rejecting local realism," Phys. Rev. A **84**, 062118 (2011).

[22] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," IEEE Trans. Inf. Theory **55**, 4337–4347 (2009).

[23] G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk, "Test martingales, Bayes factors and p-values," Statistical Science **26**, 84–101 (2011).

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

[25] 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).

[26] Tommaso Lunghi, Jonatan Bohr Brask, Charles Ci Wen Lim, Quentin Lavigne, Joseph Bowles, Anthony Martin, Hugo Zbinden, and Nicolas Brunner, "Self-testing quantum random number generator," Phys. Rev. Lett. **114**, 150501 (2015).

[27] Thomas Van Himbeeck, Erik Woodhead, Nicolas J. Cerf, Raúl García-Patrón, and Stefano Pironio, "Semi-device-independent framework based on natural physical assumptions," Quantum **1**, 33 (2017).

[28] E. Knill, Y. Zhang, and P. Bierhorst, "Quantum randomness generation by probability estimation with classical side information," (2017), arXiv:1709.06159.

[29] "See Supplemental Material at http://link.aps.org/supplemental/x.xx/physrevaxx.xxxx, which includes Refs. [40–52] for details.".

[30] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," IEEE Trans. Inf. Theory **55**, 5840–5847 (2009).

[31] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," Phys. Rev. Lett. **23**, 880–884 (1969).

[32] Antonio Acín, Serge Massar, and Stefano Pironio, "Randomness versus nonlocality and entanglement," Phys. Rev. Lett. **108**, 100402 (2012).

[33] O. Nieto-Silleras, S. Pironio, and J. Silman, "Using complete measurement statistics for optimal device-independent randomness evaluation," New J. Phys. **16**, 013035 (2014).

[34] J.-D. Bancal, L. Sheridan, and V. Scarani, "More randomness from the same data," New J. Phys. **16**, 033011 (2014).

[35] P. H. Eberhard, "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment," Phys. Rev. A **47**, R747–R750 (1993).

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

[37] Antonio Acín, Richard Gill, and Nicolas Gisin, "Optimal Bell tests do not require maximally entangled states," Phys. Rev. Lett. **95**, 210402 (2005).

[38] Yanbao Zhang, Emanuel Knill, and Scott Glancy, "Statistical strength of experiments to reject local realism with photon pairs and inefficient detectors," Phys. Rev. A **81**, 032117 (2010).

[39] B. S. Cirelśon, "Quantum generalizations of Bell's inequality," Lett. Math. Phys. **4**, 93 (1980).

[40] Y. Zhang, S. Glancy, and E. Knill, "Efficient quantification of experimental evidence against local realism," Phys. Rev. A **88**, 052119 (2013).

[41] MC Pardo and Igor Vajda, "About distances of discrete distributions satisfying the data processing theorem of

information theory," IEEE Trans. Inf. Theory **43**, 1288–1293 (1997).

[42] J. Ville, *Etude Critique de la Notion de Collectif* (Gauthier-Villars, Paris, 1939).

[43] B. G. Christensen, A. Hill, P. G. Kwiat, E. Knill, S. W. Nam, K. Coakley, S. Glancy, L. K. Shalm, and Y. Zhang, "Analysis of coincidence-time loopholes in experimental Bell tests," Phys. Rev. A **92**, 032130 (2015).

[44] Jun Shao, *Mathematical Statistics*, 2nd ed. (Springer, New York, 2003).

[45] R. König and B. Terhal, "The bounded-storage model in the presence of a quantum adversary," IEEE Trans. Inf. Theory **54**, 749–62 (2008).

[46] E. Knill, S. Glancy, S. W. Nam, K. Coakley, and Y. Zhang, "Bell inequalities for continuously emitting sources," Phys. Rev. A **91**, 032105 (2015).

[47] S. Kullback and R. A. Leibler, "On information and sufficiency," Ann. Math. Statist. **22**, 79 (1951).

[48] L. Trevisan, "Extractors and pseudorandom generators," Journal of the ACM **48**, 860–79 (2001).

[49] W. Mauerer, C. Portmann, and V. B. Scholz, "A modular framework for randomness extraction based on trevisan's construction," (2012), arXiv:1212.0520, code available on `github`.

[50] P. Bierhorst, "Geometric decompositions of Bell polytopes with practical applications," J. Phys. A: Math. Theor. **49**, 215301 (2016).

[51] 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 (2007).

[52] 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).