

Optimization of experimental quantum randomness expansion

Amelie Piveteau¹, Alban Seguinard¹, Piotr Mironowicz^{2,1,3,*} and Mohamed Bourennane¹

¹Department of Physics, *Stockholm University*, S-10691 Stockholm, Sweden

²Center for Theoretical Physics, *Polish Academy of Sciences*, Aleja Lotników 32/46, 02-668 Warsaw, Poland

³Department of Algorithms and System Modeling, Faculty of Electronics, Telecommunications and Informatics, *Gdańsk University of Technology*, Poland

 (Received 7 November 2024; revised 28 July 2025; accepted 12 August 2025; published 30 October 2025)

Quantum technologies provide many applications for information processing tasks that are impossible to realize within classical physics. These capabilities include such fundamental resources as generating secure, i.e., private and unpredictable random values. Yet, the problem of quantifying the amount of generated randomness is still not fully solved. This work presents a comprehensive analysis of the design and performance optimization of a quantum random number generator (QRNG) based on Bell inequality violations. We investigate key protocol parameters, including the smoothing parameter (ϵ_s), test round probability (γ), and switching delays, and their effects on the generation rate and quality of randomness. We identify optimal ranges for γ and p_Ω (the protocol's nonaborting probability) to balance the trade-off between randomness consumption and net randomness generation. We investigate the impact of the laser power on the overall generation rate. We explore the impact of switching delays on the system's performance, providing strategies to mitigate these effects. The work provides practical guidelines for the efficient and secure design of QRNG systems and other cryptographic protocols.

DOI: [10.1103/hc3f-3pbw](https://doi.org/10.1103/hc3f-3pbw)

I. INTRODUCTION

Quantum randomness certification is a crucial aspect of quantum information theory, particularly in applications such as cryptography, secure communications, and random number generation [1–4]. In classical systems, randomness can be generated using physical processes or algorithms; however, ensuring that this randomness is truly unpredictable and not biased or manipulated is challenging, especially when the devices are not trusted [5]. Quantum mechanics offers a unique approach to randomness through phenomena like superposition and entanglement [6].

Device-independent (DI) quantum protocols are designed to operate without trusting the internal workings of the devices used for generating random numbers [7,8]. This means that even if the devices are potentially flawed or compromised, one can still certify the randomness of the outcomes based on observed correlations between the results of measurements made on entangled quantum states. The entropy accumulation theorem (EAT) provides

a framework for quantifying how much randomness can be certified from a sequence of experiments performed in a DI manner [9,10]. The EAT theorem outlines how subsequent measurements can contribute to the overall entropy when certain conditions are met, primarily focusing on correlations that exceed classical limits often quantified using Bell inequalities [11,12].

The paper is organized as follows. In Sec. II, we present a selection of Bell inequalities used as certificates of randomness. We explore the criteria for choosing appropriate Bell expressions to ensure robust randomness certification, which is fundamental to the operation of quantum random number generators (QRNGs). Section III delves into EAT and the certification parameters used in our analysis. We detail the theoretical framework that underpins the certification of randomness, highlighting key parameters such as the smoothing parameter (ϵ_s) and its implications for randomness quality. In Sec. IV, we describe the experimental setup used to implement the QRNG protocol. We provide a detailed account of the components and configuration, including the sources of entangled photons and detection mechanisms. Section V focuses on adjusting pump laser strength to optimize the generation rate of random numbers. We analyze the trade-offs between event rates and the quality of Bell violations, providing insights into the optimal operational conditions for the QRNG.

*Contact author: piotr.mironowicz@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by [Bibsam](https://www.bibsam.org/).

The following sections focus on analysis of the possibility of fine tuning the QRNG parameters. In Sec. VI, we examine the impact of the smoothing parameter ϵ_s and generation time on the net generation rate. We discuss how different levels of the smoothing parameter affect the stability and reliability of the protocol, and the practical considerations for selecting an appropriate generation time. Section VII investigates the role of confidence intervals and test rounds in the QRNG protocol. We explain the necessity of balancing the number, or probability, of test rounds γ , and the confidence level, or the probability that the protocol does not abort (p_Ω), to ensure the robustness of the randomness certification process. In Sec. VIII, we explore the impact of switch delay on the generation rate when changing measurement settings. We analyze how delays in the experimental setup can affect the efficiency of the QRNG and propose strategies to mitigate these effects. Finally, in Sec. IX, we provide a discussion and conclusions. We summarize the findings of our study, highlighting the key factors influencing the performance and security of QRNGs. We also suggest potential avenues for future research and improvements in QRNG technology.

II. BELL CERTIFICATES OF RANDOMNESS

Bell inequalities serve as a test for the predictions of quantum mechanics against those of classical physics, particularly local hidden variable (LHV) theories. They were derived by John Bell in 1964 and are designed to assess whether the correlations observed between measurements on entangled particles can be explained by any deterministic theory that adheres to locality [6,11]. The Tsirelson bound refers to the upper limit on the strength of correlations allowed by quantum mechanics when measuring entangled particles [13]. While classical theories can reach at most a certain limit, the classical bound, quantum mechanics allows for stronger correlations than those permitted by classical physics but still respects this Tsirelson bound. The significance of the bounds lie in their role in distinguishing between different types of correlations, those achievable through classical means and those achievable through quantum entanglement. If Bell inequality is violated it gives an opportunity to provide certification of security of random numbers even when devices are not trusted [1]. There is on going research on finding the most suitable Bell inequalities for this purpose [14,15].

We define the correlators $\text{Cor}(x,y)$ between Alice's results of the measurement x and Bob's results of the measurement y with the following equation:

$$\begin{aligned} \text{Cor}(x,y) \equiv & P(0,0|x,y) + P(1,1|x,y) \\ & - P(0,1|x,y) - P(1,0|x,y). \end{aligned} \quad (1)$$

The CHSH inequality was formulated by Clauser, Horne, Shimony, and Holt in 1969 [16]. It involves two parties (Alice and Bob) each choosing between two measurement settings (A0, A1 for Alice and B0, B1 for Bob) and is defined as $\text{CHSH} \equiv \text{Cor}(0,0) + \text{Cor}(0,1) + \text{Cor}(1,0) - \text{Cor}(1,1)$.

Other certificates belong to a family introduced recently in Ref. [15]. The family is parameterized by $\delta \in (0, \pi/6]$, and is defined as

$$\begin{aligned} I_\delta \equiv & \text{Cor}(0,0) \\ & + \frac{1}{\sin \delta} (\text{Cor}(0,1) + \text{Cor}(1,0)) - \frac{1}{\cos 2\delta} \text{Cor}(1,1). \end{aligned} \quad (2)$$

Members of this family exhibit self-testing properties, use two measurement settings per party, and can certify two bits of randomness for the settings $x^* = y^* = 0$.

The work [15] proposed also a second family, parameterized by $\gamma \in [0, \pi/12]$, which defines the Bell expressions:

$$\begin{aligned} J_\gamma \equiv & \text{Cor}(0,0) + (4 \cos^2 [\gamma + \pi/6] - 1) \\ & \times (\text{Cor}(0,1) + \text{Cor}(1,0) - \text{Cor}(1,1)). \end{aligned} \quad (3)$$

These expressions also use two settings and exhibit self-testing properties, although in most cases they do not certify two bits of randomness. Setting $\gamma = \pi/12$ in Eq. (3) yields the standard CHSH expression.

The Tsirelson bounds for Eqs. (2) and (3) are given by

$$I_\delta^Q \equiv \frac{2 \cos^3 \delta}{\cos(2\delta) \sin \delta}, \quad J_\gamma^Q \equiv 8 \cos^3 (\gamma + \pi/6). \quad (4)$$

Their classical bounds are

$$I_\delta^L = -1 + \frac{2}{\sin \delta} + \frac{1}{\cos 2\delta}, \quad J_\gamma^L = 12 \cos^2 \left(\gamma + \frac{\pi}{6} \right) - 4. \quad (5)$$

For the family, Eq. (2), we selected $\delta = 0.52$ as the main focus of this paper. As shown in previous studies [17], as δ approaches $\pi/6 \approx 0.5236$, the amount of certified randomness increases.

To certify the randomness, we utilized two Bell expressions, viz. CHSH, and one from a family introduced in Ref. [15] and defined with the following expression:

$$\begin{aligned} & \text{Cor}(0,0) + 2.0126 \times \text{Cor}(0,1) \\ & + 2.0126 \times \text{Cor}(1,0) - 1.9754 \times \text{Cor}(1,1). \end{aligned} \quad (6)$$

From Eq. (5) it follows that the local bound for $\delta = 0.52$ is 5.0005.

It was shown in Ref. [15] that, up to local isometries, there exists a unique quantum strategy that achieves the

maximal quantum value of the Bell expression, i.e., $I_\delta = I_\delta^Q$. The optimal state used in this strategy is the maximally entangled two-qubit state

$$\rho_{QAOB} = |\psi\rangle\langle\psi|, \quad \text{where } |\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (7)$$

The measurements are defined as follows. Alice's observables are

$$A_0 = \sigma_Z, \quad A_1 = -\sin \delta \sigma_Z + \cos \delta \sigma_X, \quad (8)$$

while Bob's observables are

$$B_0 = \sigma_X, \quad B_1 = \cos \delta \sigma_Z - \sin \delta \sigma_X, \quad (9)$$

where σ_Z and σ_X denote the standard Pauli operators. The parameter δ defines the Bell expression I_δ , and the choice of measurements depends explicitly on this value.

Although in this paper we focus on two specific Bell expressions, including the standard CHSH inequality, the results can be easily extended to other QRNG protocols. We analyze the influence of various parameters of the EAT on performance, and similar insights into the role of these parameters will be useful for designers of future protocols. The focus on CHSH stems from its simplicity and popularity, which makes the community particularly interested in seeing that, even in such a basic case, the choice of protocol parameters is not straightforward and can significantly affect the generation and expansion rates.

III. ENTROPY ACCUMULATION AND CERTIFICATION PARAMETERS

We utilize these two Bell expressions for quantum randomness certification with various parameters. In the work we follow the formulation of EAT provided in Refs. [18–20]. A crucial tool needed to calculate the certified entropy with EAT is the min-tradeoff function, providing a linear lower bound on the entropy in a single event for a given value of Bell violation.

A. Entropy accumulation framework

To analyze finite-size effects in our protocol, we use the EAT [18,19]. This framework models a sequence of rounds as *EAT channels* $\{\mathcal{N}_i\}_{i=1}^n$, where each channel is a completely positive trace-preserving map:

$$\mathcal{N}_i : \mathcal{S}(R_{i-1}) \longrightarrow \mathcal{S}(A_i B_i X_i Y_i C_i R_i), \quad (10)$$

with classical outputs A_i, B_i, X_i, Y_i, C_i (where C_i is a deterministic function of the others) and an arbitrary quantum

register R_i . The EAT assumes that for any initial state ρ_{R_0E} , the final state

$$\rho_{A^n B^n X^n Y^n C^n E} := \text{Tr}_{R_n} [(\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1 \otimes \text{id}_E)(\rho_{R_0E})] \quad (11)$$

satisfies, for each i , the *Markov condition*:

$$I(A^{i-1} B^{i-1} : X_i Y_i | X^{i-1} Y^{i-1} E) = 0. \quad (12)$$

This condition ensures that inputs (X_i, Y_i) are chosen independently of previous outcomes and adversarial side information, given past inputs.

To apply EAT, one defines a *min-tradeoff function* $f : \mathcal{P}_C \rightarrow \mathbb{R}$, where $\mathcal{P}_C$ is the set of distributions over the classical register C_i . The function f must satisfy

$$f(p) \leq \inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i(\sigma)_{C_i} = \tau_p} H(A_i B_i | X_i Y_i R')_{\mathcal{N}_i(\sigma)}, \quad (13)$$

for any $p \in \mathcal{P}_C$, where $\tau_p = \sum_{c \in \mathcal{C}} p(c) |c\rangle\langle c|$. In practice, such functions are constructed using SDP relaxations (e.g., NPA [21,22] or BFF [23]) that bound the conditional entropy given observed statistics.

Theorem 1 (Entropy accumulation theorem [18,19]).

Let $\rho_{A^n B^n X^n Y^n C^n E}$ be the global state after applying the EAT channels $\{\mathcal{N}_i\}_{i=1}^n$ to some ρ_{R_0E} , and let $\Omega \subseteq \mathcal{C}^n$ be an event on the classical outputs with probability $p_\Omega = \Pr[C^n \in \Omega]$. Let $\rho_{|\Omega}$ be the postselected state on this event. Suppose a valid min-tradeoff function f satisfies

$$f(\text{freq}(C^n)) \geq t \quad (14)$$

for all $C^n \in \Omega$. Then for any smoothing parameter $\epsilon_s \in (0, 1)$ and $\beta \in (0, 1)$, the smooth min-entropy satisfies

$$H_{\min}^{\epsilon_s}(A^n B^n | X^n Y^n E)_{\rho_{|\Omega}} > nt - n(\epsilon_V + \epsilon_K) - \epsilon_\Omega, \quad (15)$$

with the correction terms defined as

$$\epsilon_V := \frac{\beta \ln 2}{2} \left(\log(2|\mathcal{AB}|^2 + 1) + \sqrt{\text{Var}[f |_\Gamma] + 2} \right)^2, \quad (16a)$$

$$\epsilon_K := \theta_1 \times \theta_2 \times \theta_3, \quad (16b)$$

$$\epsilon_\Omega := \frac{1}{\beta} (1 - 2 \log(p_\Omega \epsilon_s)), \quad (16c)$$

where

$$\theta_1 := \frac{\beta^2}{6(1 - \beta)^3 \ln 2}, \quad (17a)$$

$$\theta_2 := 2^{\beta(\log |\mathcal{AB}| + d_f)}, \quad (17b)$$

$$\theta_3 := \ln^3(2^{\log |\mathcal{AB}| + d_f} + e^2), \quad (17c)$$

$$d_f := \max f - \min f |_\Gamma. \quad (17d)$$

This bound shows that, conditioned on Ω , one accumulates at least nt bits of entropy up to finite-size penalties. The terms ϵ_V , ϵ_K , and ϵ_Ω capture statistical variance, higher-order corrections, and smoothing cost, respectively. In practice, C_i encodes Bell inequality statistics; the trade-off function f is obtained via numerical optimization; and the min-entropy bound (15) determines the number of extractable random bits or key rate per round.

The quantity $\text{Var}[f|_\Gamma]$ denotes the worst-case variance of $f(C_i)$ conditioned on Ω , and $d_f = \max f - \min f|_\Gamma$ captures the range of f over relevant distributions. The parameters $|\mathcal{AB}|$ refer to the size of the classical output alphabets per round. For precise definitions and a detailed derivation of the bound, we refer the reader to the formal framework developed in Ref. [19].

B. Parameters of EAT

The first considered parameter is the so-called smoothing parameter of randomness, ϵ_s . This parameter specifies our requirement about the quality of the randomness, as further used by randomness extractors. The smaller the value of ϵ_s the higher is the amount of certified randomness.

The other parameter is the value of the violation of the Bell inequality. The higher the quality of the experiment, the higher the violation should be. On the other hand, in the experiment using the spontaneous parametric down-conversion (SPDC) process, a higher emission rate of entangled pairs of photons usually decreases the quality of the particular Bell violation, as discussed in Ref. [24]. The number of detected pairs depends on the efficiency of detectors and transmission and coupling losses, and its rate is denoted as *events per second*.

Any real-world device must produce results within a finite time to be practically useful. By *generation time*, we mean the total duration of a single certified randomness generation session on the quantum part of the QRNG. In our scheme, each generation consists of a sequence of rounds, which are as either test and generation, as described below. Both types are essential and interleaved throughout the process. To obtain quantum data for later extraction of certified randomness, the user performs a sequence of such rounds during a single experimental run. Its total duration is what we refer to as the *generation time*.

This duration does not include classical operations such as postprocessing or randomness extraction. Finite-size effects are accounted for using the EAT, which requires setting a generation time to determine the number of events available for statistical analysis. Knowing both the generation time and the event rate allows us to compute the total number of events (as their product). Longer generation times reduce the negative impact of finite-size corrections in EAT, improving randomness yield. However, shorter generation times are desirable for practical devices, as the

user receives randomness in chunks corresponding to full generation sessions and typically does not want to wait too long for usable output.

For randomness generation, only a particular pair of settings (x_0, y_0) of Alice and Bob are considered. To estimate the value of Bell violation in some rounds of the protocol, we need to randomly choose settings different from (x_0, y_0) . The rounds used for generation are called *generation rounds*, and those used for testing are called *test rounds*. The probability of test rounds is denoted by γ . This approach is called spot checking [25]. Every test round consumes public randomness for the random choice of settings x and y ; generation rounds do not consume additional randomness. The larger γ , the better the estimation of Bell violation, but more randomness is consumed, and fewer rounds are used for generating new randomness. In certification, we know the Bell violation only up to a given confidence interval, influencing the probability that the protocol does not abort, p_Ω , and we always need to assume the most pessimistic violation value to keep the protocol secure. The more events are used for Bell value estimation, the lower is the risk that the value fails out of the assumed interval by chance. Therefore, there is a nontrivial relation between the generation rate and the values of γ and p_Ω . Each of the generation rounds is producing certain entropy $H(AB|x_0, y_0, E)$ for outputs A and B and quantum side information E .

Since we need to change the optical components settings in the experimental setup for the test rounds, every change of settings usually introduces a certain delay, depending on the devices used. This delay additionally decreases the rate of events and, consequently, the rate of the protocol.

The parameters discussed above directly relate to the devices' physical workings. Apart from them, the formulae appearing in EAT contain an additional parameter, usually denoted as β , which is in the range $(0, 1)$ and can be chosen arbitrarily.

The EAT analysis was performed using the `expdqrng` framework [20].

IV. EXPERIMENTAL SETUP

The experimental setup is illustrated in Fig. 1. To generate the polarization entangled pair in state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$, two 2-mm-thick β barium borate nonlinear crystals type 1 placed in interferometric configuration are pumped with a femtosecond in pulsed laser at a wavelength of 390 nm, to produce photon pairs emitted into two spatial modes (a) and (b) through the degenerate SPDC process. The spatial, spectral, and temporal distinguishability between the down-converted photons is carefully removed by coupling to single-mode fiber, passed through narrow-bandwidth interference filters (F) and quartz wedges, respectively. The pump power of the femtosecond laser is tunable, allowing the photon-pair

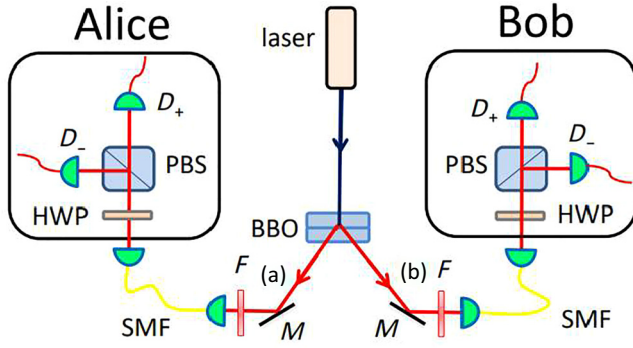


FIG. 1. Experimental setup. Entangled photon pairs are generated through the SPDC process. The emitted photons in modes (a),(b) are coupled to single-mode fiber (SMF) and pass narrow filters (F). Each of the two stations' measurements is composed of a half-wave plate (HWP), a polarization beam splitter (PBS), and single photon detectors (D_+ and D_-). (See main text for details.)

generation rate to be set according to the experiment's requirements. Alice and Bob's measurements are performed with the help of half-wave plates (HWP_A , HWP_B) oriented, respectively, θ_{A0} or θ_{A1} , and θ_{B0} or θ_{B1} , where θ_{A0} , θ_{A1} , θ_{B0} , and θ_{B1} are the angles achieving the Tsirelson bound to each of the considered Bell expression. The polarization measurement was performed using PBS, and single-photon detectors (D) were placed in the two output modes of the PBS. Our detectors are actively quenched Siavalanche photodiodes. All single-detection events were registered using a VHDL-programmed multichannel coincidence logic unit with a time-coincidence window of 1.7 ns. In this proof-of-principle experiment, we have not used real-time switching. Instead the measurements were performed in series, for a certain time, and after each series we switched for next measurement settings.

The photo pair emission rate was set for each measurement, and the corresponding measurement time was adjusted to keep the same number of two-photon coincidence events for the measurement setting. At low rates, the multiphoton pair emissions are small, and therefore, the accidental events can be neglected. At higher rates, the

accidental events will affect the correlations and the level of inequality violation.

V. ADJUSTING SPDC PUMP LASER POWER FOR GENERATION RATE

The first decision when designing a QRNG is selecting the appropriate Bell inequality to use as the certificate. Next, we must determine how to adjust the SPDC pump laser power to achieve the optimal generation rate. This involves balancing the number of events per second with the Bell violation, as a higher laser power typically results in more events per second but a lower Bell violation, leading to less certified randomness per event. To estimate the optimal settings, we tested various event rates for the two Bell inequalities, CHSH, and Eq. (6). We then calculated the asymptotic randomness generation rate for an infinite number of repetitions, thereby ignoring EAT's finite data size effects.

A. Asymptotic randomness generation rates

We computed the amount of certified entropy using two methods: the Navascués-Pironio-Acín (NPA) method [21,22], which calculates a lower bound on certified min-entropy [26–29], and the Brown-Fawzi-Fawzi (BFF) method [30], which calculates a lower bound on certified von Neumann entropy. The numerical calculation were conducted using NCPOL2SDPA [31], MOSEK Solver [32].

Both NPA and BFF can be executed at various levels, forming a hierarchy. Higher levels offer better approximations but require more computational resources. We used level 2 of the hierarchy for NPA, which needed about a dozen seconds of computation. Higher levels, such as 3 and 4, did not improve the approximation significantly in our cases. For BFF, we also used level 2 of the hierarchy. Additionally, BFF involves another parameter: the number of nodes in the Gauss-Radau quadrature used in the approximation formulas. We considered using six and eight nodes. The computations took approximately 2.5 min for Eq. (6) and 7 min for CHSH with six nodes. With eight nodes, the computation times increased to 3 min for

TABLE I. Dependence of the certified min-entropy ($H_{\min}$) and von Neumann entropy (vNe) on the violation of the Bell inequality (6) and the number of events per second. The values of $H_{\min}$ are calculated with NPA level 2. The values of von Neumann entropy are calculated with approximations using Gauss-Radau quadrature with six or eight nodes. The value 4.95151 is below the local bound 5.0005, and thus does not certify any randomness.

Events per second	Bell value	$H_{\min}$	vNe Radau 6	vNe Radau 8	Asymptotic rate
28 000	4.95151	0	0	0	0
24 000	5.00247	0.0098	0.0186	0.0186	446
20 000	5.02311	0.1231	0.2234	0.2244	4488
16 000	5.03036	0.1651	0.2953	0.2965	4744
12 000	5.04098	0.2289	0.4007	0.4024	4829
8000	5.08671	0.5413	0.8545	0.858	6864

TABLE II. Dependence of the certified min-entropy ($H_{\min}$) and von Neumann entropy (vNe) on the violation of the Bell inequality CHSH and the number of events per second. The values of $H_{\min}$ are calculated with NPA level 2. The values of von Neumann entropy are calculated with approximations using Gauss-Radau quadrature with six or eight nodes. i is the experiment number, as discussed in Sec. VB, where additional data are provided in Table III.

i	Events per second	Bell value	$H_{\min}$	vNe Radau 6	vNe Radau 8	Asymptotic rate
1	70 000	2.6502	0.5198	0.8909	0.8964	62 748
2	52 000	2.6760	0.5638	0.9497	0.9574	49 785
3	36 000	2.7026	0.6148	1.0156	1.0239	36 860
4	20 000	2.7150	0.6411	1.0483	1.0566	21 132
5	12 000	2.7369	0.6925	1.1092	1.1177	13 412
6	8000	2.7443	0.7117	1.1309	1.1397	9118
7	4000	2.7609	0.7591	1.1831	1.1917	4767

Eq. (6) and 7.5 min for CHSH, but the resulting estimates improved by only about 0.4% and 0.8%, respectively. The min-tradeoff function is obtained from the dual solutions of semi-definite optimization used in NPA and BFF with the method from Sec. 3.4 of Ref. [23]. The min-tradeoff function has to be calculated once for each Bell violation value and does not need to be recalculated in further fine tuning of protocol parameters.

The results are shown in Table I for Eq. (6) and Table II for CHSH. Table I shows that the best asymptotic rate for Eq. (6) is achieved at 8000 events per second. Deviating from this event rate results in lower rates. Conversely, as shown in Table II, the optimal asymptotic rate for CHSH occurs at the highest considered event rate, 70 000 events per second. We expect that further increasing the event rate will eventually cause a similar effect as seen in Table I, where increasing the pump laser power beyond a certain point will start to decrease the asymptotic generation rate. We provide the extrapolation of the results in Sec. VB using the techniques introduced in Ref. [24]. For the remainder of this work, in the sections following Sec. VI, we will focus on the best asymptotic case, 70 000 events per second for CHSH. Note, that since these tables are concerning the asymptotic case, the choice of the value of the parameter γ is irrelevant.

TABLE III. Single and coincidence detection counts, denoted by $c_{\text{obs}}^{(i)}$ and $s_{\text{obs}}^{(i)}$, respectively, are recorded alongside visibility measurements for CHSH at different times of collecting events $t^{(i)}$. Each experimental run corresponds to the scenarios listed in Table II.

i	$c_{\text{obs}}^{(i)}$	$s_{\text{obs}}^{(i)}$	$t^{(i)}$
1	37 892 989	549 605 351	540
2	43 322 946	660 223 194	832
3	40 639 747	646 698 789	1112
4	40 767 056	631 786 125	2000
5	41 386 494	668 668 812	3332
6	40 381 162	650 423 503	5000
7	36 888 729	590 756 887	10 000

B. Behavior for increasing pump laser power

Table III presents raw data collected during the experiment. Here, $c_{\text{obs}}^{(i)}$ and $s_{\text{obs}}^{(i)}$ are the observed number of double and single clicks, and $t^{(i)}$ is the measurement time (in seconds) for i th experimental case. $B_{\text{obs}}^{(i)}$, the observed Bell values, are provided in Table II.

In order to extrapolate the Bell value, denoted B_{calc} , and estimate relevant events per second we used the method of Ref. [24]. With the predicted values we calculate the asymptotic rates. The peak performance is predicted to be achieved around Bell value 2.3 at about 470 588 events per second. The results are shown in Table IV. Nonetheless in the experiment we were able to achieve at most 70 000 events per second. This limit stems from the power limit of the pump laser.

VI. SMOOTHING PARAMETER AND GENERATION TIME

As discussed earlier, one crucial parameter in QRNG is the smoothing parameter, ϵ_s . The smaller this value, the higher the quality of the generated randomness, which is crucial for various applications. In our research, we explored different values of the smoothing parameter and

TABLE IV. Extrapolated values of asymptotic generation rates when increasing the number of events per second using CHSH as the certificate. The maximal asymptotic generation rate is for the value of CHSH equal 2.3. B_{calc} is obtained using the method of Ref. [24] calculated from the experimental data from Tables II and III.

Events per second	B_{calc}	vNe Radau 8	Asymptotic rate
93 240	2.625	0.8406	78 375
113 636	2.6	0.7887	89 628
207 254	2.5	0.6050	125 397
322 581	2.4	0.4469	144 160
470 588	2.3	0.3088	145 322
655 738	2.2	0.1881	123 334
888 889	2.1	0.0844	75 007
1 212 121	2	0	0

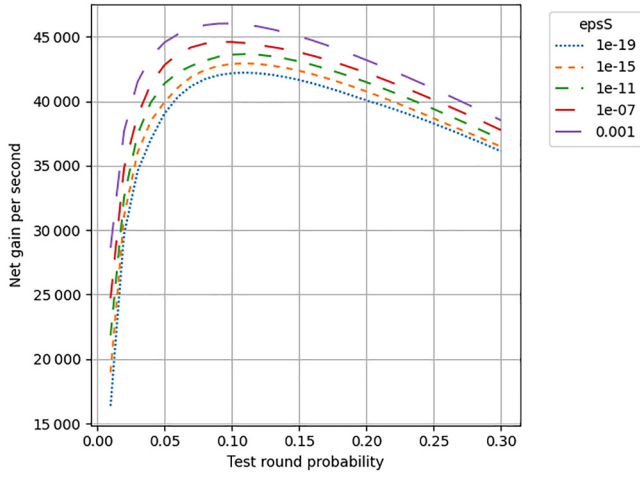


FIG. 2. The net generation rate for 70 000 events per second using the Bell inequality CHSH for various values of the smoothing parameter ϵ_s , and test round probability γ for generation time of 600 s. The value of ϵ_s changes the generation rate by at most around 10%, even though we consider a wide range of values from 10^{-3} to 10^{-19} . This indicates that the impact of ϵ_s on the performance is marginal in practice.

observed that for the investigated protocol, reducing the smoothing parameter only slightly decreases the generation rate, as shown in Fig. 2. Therefore, we opted for a much higher quality of randomness with $\epsilon_s = 1 \times 10^{-15}$. On the other hand, as discussed further, a proper selection of the smoothing parameter can have more significant impact, see further Fig. 8.

Another significant parameter for practical applications of QRNG is generation time. Users typically prefer randomness to be generated as quickly as possible. However, while immediate generation is ideal, it is feasible to store larger segments, referred to as *single data chunks*, of generated randomness and supply them on demand. This means that the generation time does not necessarily need to be in the milliseconds range; a time frame of several minutes is often acceptable.

The relationship between the net generation rate and the generation time for a single data chunk is illustrated in Fig. 3. The data indicates that a minimum of 6 min is required to achieve a positive rate. The rate increases rapidly for generation times under 1 h, reaching satisfactory levels at around 10 min. Consequently, we focus on these two generation times in our subsequent analysis. The data point at 19.2 h demonstrates that our proposed protocol achieves over 50 kbps, significantly surpassing the 13 527 bits per second achieved with the same generation time in Ref. [4].

The value of the test-round probability γ at each point in Fig. 3 was optimized for each data chunk separately and can vary significantly depending on the generation time. For example, as follows:

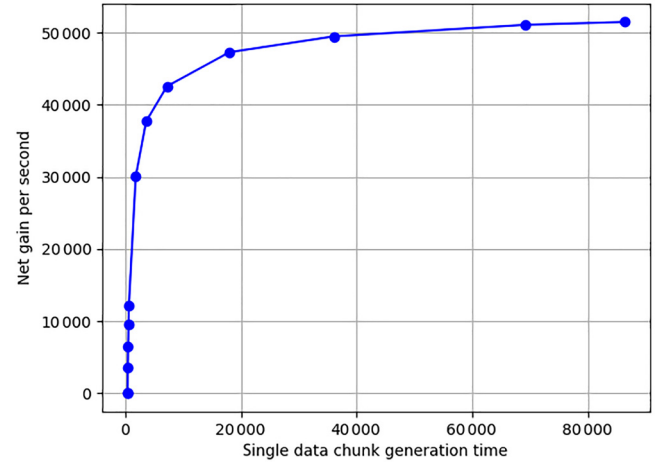


FIG. 3. The net generation rate for 70 000 events per second using the Bell inequality CHSH, with the smoothing parameter $\epsilon_s = 1 \times 10^{-15}$, depends significantly on the generation time for each data chunk. The minimum generation time required to achieve a positive rate is 6 min. The range of generation times explored extends up to 24 h. It is worthwhile to note that the asymptotic generation rate reported in Table II exceeds the values shown in the plot. This suggests that increasing the generation time can further mitigate the impact of finite data size effects, thereby enhancing the certified randomness rate. The data point at 69 120 s (19.2 h) was specifically chosen to compare with the results presented in Ref. [4], where this value was utilized.

- (a) For a generation time of 420 s: the optimal $\gamma = 0.01$.
- (b) For 600 s: $\gamma = 0.1$.
- (c) For 86 400 s (i.e., 24 h): $\gamma = 0.001$.

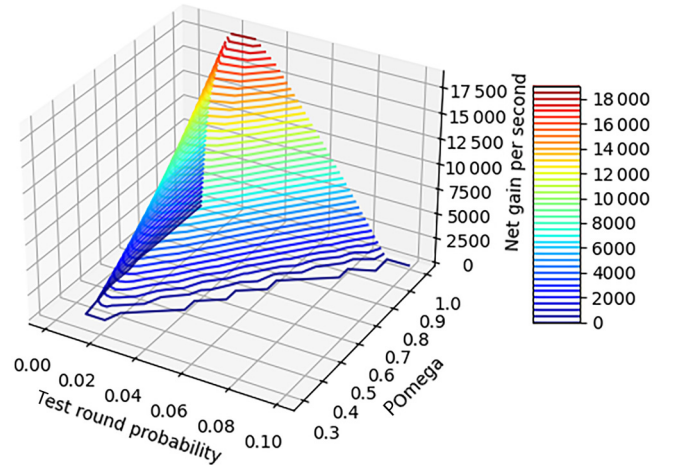


FIG. 4. Effect of the test round probability γ and the protocol nonaborting probability p_Ω on the generation rate, for a single data chunk generation time of 10 min. This analysis assumes the smoothing parameter $\epsilon_s = 1 \times 10^{-15}$ and uses the Bell inequality CHSH at an event rate of 70 000 per second.

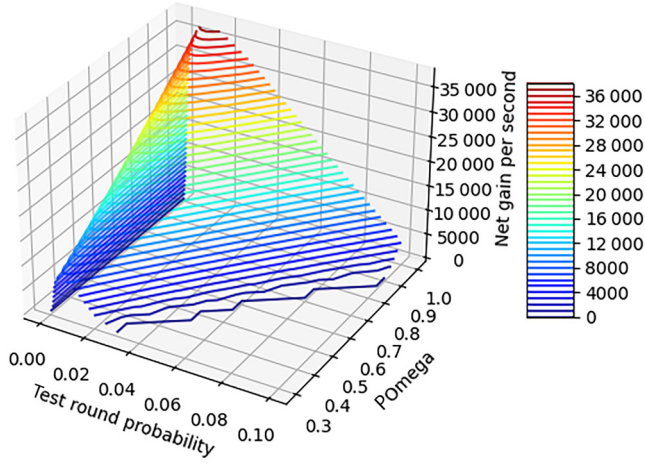


FIG. 5. Impact of the test round probability γ and the protocol's nonaborting probability p_Ω on the generation rate, with a single data chunk generation time of 1 h. The protocol parameters are consistent with those used in Fig. 4.

As evident, the dependency is not monotonic. The longer the total time, the more samples are available for estimating the Bell expression, so test rounds can, in principle, occur less frequently. On the other hand, the more frequent the test rounds, the higher the overall cost of randomness used for setting choice. When the generation per round is low, frequent testing may not be compensated by sufficient randomness expansion in the generation rounds. Therefore, it may be beneficial to slightly compromise the accuracy of the Bell expression estimation, accepting a lower generation per round, in order to reduce the burden of setting choice randomness. This trade-off is clearly illustrated in Fig. 2, which shows an optimum around $\gamma = 0.1$.

VII. CONFIDENCE INTERVALS AND TEST ROUNDS

The next two critical elements in tuning a QRNG are the confidence intervals, expressing the probability that the protocol does not abort (denoted as p_Ω), and the probability of test rounds (denoted as γ). Even if the average generation rate is high, a lower p_Ω means the generation process is less stable. In other words, if p_Ω is low, there is a relatively high probability that the Bell violation might not be significant enough to accept the generated chunk of randomness after the entire generation procedure. This can be a major drawback for the user experience, as the user might have to wait for the data to find out it cannot be used. To reduce the likelihood of the protocol aborting, one can either increase the tolerance for lower Bell violations or conduct more test rounds, thereby mitigating the risk that a lower Bell violation is due to statistical fluctuations rather than an eavesdropper's influence.

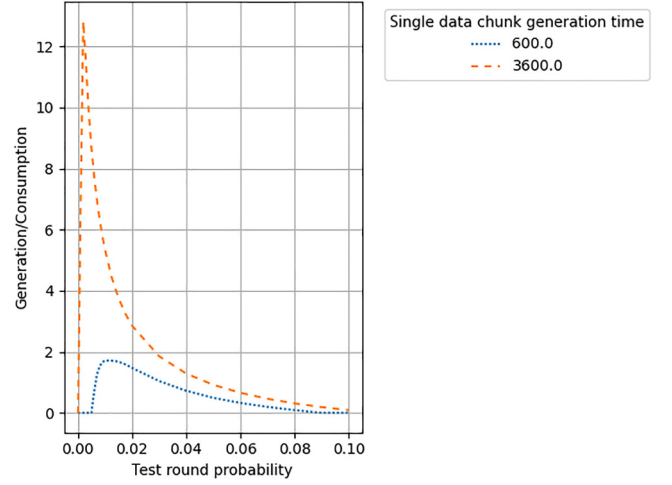


FIG. 6. The expansion rate, defined as the ratio of total net randomness generated to the total randomness consumed, for single data chunk generation times of 10 min and 1 h. The analysis considers the smoothing parameter $\epsilon_s = 1 \times 10^{-15}$ and employs the Bell inequality CHSH with an event rate of 70 000 per second.

Thus, while test rounds can help stabilize the protocol's operation, they come with certain additional costs. The most explicit cost is the increased consumption of randomness required for these rounds. It is worthwhile to note that test rounds fail to generate new randomness and require additional randomness (2 bits for two settings) to choose the settings of Alice and Bob for estimating the Bell violation. Furthermore, selecting which rounds are test rounds also consumes randomness, although this is usually much smaller than the randomness needed during the test rounds themselves.

Alice and Bob's settings are chosen according to probability distributions P_X and P_Y , respectively. Let H_X and H_Y denote the Shannon entropies of P_X and P_Y . We have $H_X = H_Y = 1$ for uniform distribution of settings. The total randomness consumption for a single choice of settings for Alice and Bob in test rounds is denoted by $H_{XY} = H_X + H_Y$. The randomness consumed in selecting which rounds are test rounds is given by the binary entropy of γ , denoted as $H_2(\gamma)$. This randomness is consumed in every round, not just the test rounds. Therefore, the average consumption per round is calculated as

$$H_{XY} \times \gamma + H_2(\gamma). \quad (18)$$

The generation rate's dependence on the test round probability γ and the protocol's nonaborting probability p_Ω for single data chunk generation times of 10 min and 1 h is shown in Figs. 4 and 5, respectively. In both cases, it was optimal to use a p_Ω value close to 1. Thus, we focus on $p_\Omega = 0.9999$ for further analysis. To estimate the corresponding tolerance for the deviation of the observed Bell violation from the expected value, we used the estimation

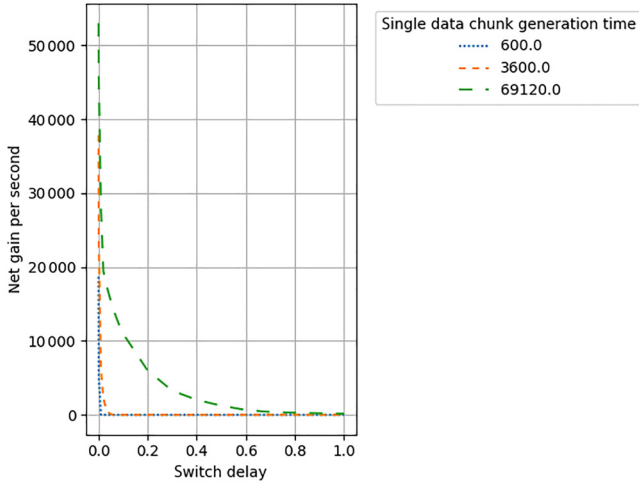


FIG. 7. Net generation rate as a function of switch delay (in seconds) for chunk generation times of 10 min, 1 h, and 19.2 h. For the 10-min chunk duration, the net gain diminishes almost immediately with any switching delay. In the case of a 1-h chunk, a small switch delay is tolerable but still results in a significant decrease in the generation rate. For a 19.2-h chunk, the rate declines gradually with increasing delay. The analysis assumes the smoothing parameter $\epsilon_s = 1 \times 10^{-15}$ and uses the Bell inequality CHSH with an event rate of 70 000 per second.

method from Ref. [1]. The plots indicate that the optimal test round probability γ decreases with longer generation times. It is worthwhile to note that this analysis assumes the test rounds' switching time is zero.

The protocol must minimize the consumption of randomness, as an additional source of secure, unpredictable randomness must be provided for its operation. While the randomness does not need to be private, it should be unknown before the protocol runs. Otherwise, untrusted devices could potentially adapt to the known randomness, exploiting deterministic strategies to simulate a higher Bell violation. A significant drawback of the approach described in Ref. [4] was its inefficiency in terms of randomness consumption. Specifically, generating 6.496×10^9 random bits required 6.233×10^9 bits, resulting in a net generation of only 2.63×10^8 bits. This led to an expansion rate of approximately $2.63 \times 10^8 / 6.233 \times 10^9 \approx 0.042$. In contrast, the present work reports a significantly better expansion rate, as demonstrated in Fig. 6.

VIII. IMPACT OF SWITCH DELAY FOR MEASUREMENT SETTINGS

The final parameter we investigate is the switching delay introduced by the time mechanical elements of the device are required to transition between various settings of Alice and Bob. Naturally, a longer switching delay has a more pronounced negative impact on the generation rate of the

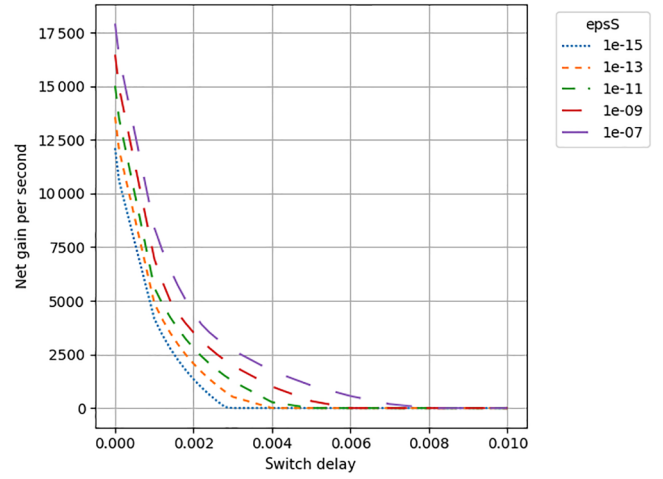


FIG. 8. Generation rate as a function of switch delay (in seconds) for a chunk generation time of 10 min and various values of the smoothing parameter ϵ_s . The plot illustrates that increasing the smoothing parameter can slightly alleviate the negative impact of longer switch delays on the generation rate. We used the Bell inequality CHSH with an event rate of 70 000 per second.

QRNG. To provide a comprehensive analysis, we consider single data chunk generation times of 10 min, 1 h, and 19.2 h, the latter included for reference as reported in Ref. [4]. The results are presented in Fig. 7.

To address the negative impact of switch delay on the generation rate, we observe that allowing a higher smoothing parameter can provide significant improvements. This effect is particularly noticeable for shorter chunk generation times. As demonstrated in Fig. 8 (cf. Fig. 2), for a chunk generation time of 10 min, the maximum switch delay that can be tolerated is 0.01s. By adjusting the smoothing parameter, we can mitigate some of the adverse effects of increased switch delays.

IX. DISCUSSION AND CONCLUSIONS

In this work, we have explored the practical considerations and performance trade-offs in designing and implementing a QRNG based on Bell-inequality violations. Specifically, we analyzed the impact of various protocol parameters, such as the smoothing parameter (ϵ_s), test round probability (γ), and switching delays, on the generation rate and randomness quality. Our findings provide insights for optimizing QRNGs in real-world applications.

The smoothing parameter is a crucial parameter in QRNGs, as it determines the quality and security of the generated randomness. A smaller ϵ_s implies a higher quality of randomness, which is critical for applications requiring high-security guarantees. However, maintaining a low smoothing parameter, such as 1×10^{-15} , is feasible and offers substantial improvements in randomness quality without severely compromising the generation rate.

Our investigation revealed that optimizing γ and p_Ω is essential for balancing the trade-off between randomness consumption and generation rate. A higher γ improves the accuracy of Bell violation estimation but reduces the net randomness generated due to increased randomness consumption in test rounds. Conversely, a high p_Ω ensures the protocol's stability by reducing the likelihood of protocol abortion, which is critical for practical usability. Our results indicate that maintaining p_Ω close to 1 is optimal, as it balances stability and efficiency.

Switching delays, which arise from the mechanical limitations of the experimental setup, significantly impact the generation rate, especially for shorter chunk generation times. Our study shows that even small delays can drastically reduce the net generation rate. However, increasing the smoothing parameter can partially compensate for the adverse effects of switching delays, providing a viable strategy for mitigating their impact.

Our results demonstrate an improvement over the QRNG design presented in Ref. [4], which had a low expansion rate and limited net randomness generation. By optimizing the protocol parameters and considering longer chunk generation times, we achieved a higher expansion rate and a more robust performance under varying conditions, including switching delays. The choice of a high-quality smoothing parameter further distinguishes our approach, providing a better balance between security and efficiency.

Future research could focus on further optimizing the QRNG protocol by exploring additional parameters and configurations. In particular, investigating alternative Bell inequalities or quantum states may yield higher generation rates or better resilience to experimental imperfections. Additionally, developing faster and more precise switching mechanisms could reduce the impact of switching delays, enhancing the QRNG's overall performance.

Randomness extractors are indispensable for converting the certified but imperfect entropy from our Bell-based protocols into nearly uniform bits suitable for cryptographic use. In particular, after certifying von Neumann entropy via device-independent Bell tests and the entropy accumulation theorem, one can apply a strong seeded extractor, such as Trevisan's extractor, to purify this entropy into uniform random bits with security against quantum adversaries [33,34]. Moreover, privacy amplification via the leftover hash lemma ensures that any residual correlations with an eavesdropper are eliminated by universal hashing [35]. By integrating these extraction techniques into the presented QRNG workflow, one can pave the way to practical implementation, yielding high-quality randomness.

Other experiments on device-independent QRNG include Ref. [36]. A closely related problem is the implementation of quantum key distribution systems, which are also based on certified entropy with Bell nonlocality

[37,38]. Another work discussing possible improvements in QRNG is Ref. [39].

Our work highlights the significance of carefully tuning QRNG protocol parameters to maximize performance and security. By considering the interplay between the smoothing parameter, test round probability, protocol stability, and switching delays, we provide a comprehensive framework for optimizing QRNGs. Our findings offer practical guidelines for designing robust and efficient QRNG systems capable of meeting the stringent requirements of secure applications.

ACKNOWLEDGMENTS

This work was supported by the Knut and Alice Wallenberg Foundation through the Wallenberg Centre for Quantum Technology (WACQT). P.M. was supported by the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101080086/NeQST.

DATA AVAILABILITY

The data are not publicly available. The data are available from the authors upon reasonable request.

-
- [1] 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, *et al.*, Random numbers certified by Bell's theorem, *Nature* **464**, 1021 (2010).
 - [2] A. Acín and L. Masanes, Certified randomness in quantum physics, *Nature* **540**, 213 (2016).
 - [3] Y. Liu, Q. Zhao, M.-H. Li, J.-Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.-Z. Liu, C. Wu, X. Yuan, *et al.*, Device-independent quantum random-number generation, *Nature* **562**, 548 (2018).
 - [4] W.-Z. Liu, M.-H. Li, S. Ragy, S.-R. Zhao, B. Bai, Y. Liu, P. J. Brown, J. Zhang, R. Colbeck, J. Fan, Q. Zhang, and J.-W. Pan, Device-independent randomness expansion against quantum side information, *Nat. Phys.* **17**, 448 (2021).
 - [5] D. Mayers and A. Yao, Self testing quantum apparatus, *Quantum Inf. Comput.* **4**, 273 (2004).
 - [6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
 - [7] R. Renner, Security of quantum key distribution, *Int. J. Quantum Inf.* **6**, 1 (2008).
 - [8] C. Portmann and R. Renner, Security in quantum cryptography, *Rev. Mod. Phys.* **94**, 025008 (2022).
 - [9] 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).
 - [10] F. Dupuis, O. Fawzi, and R. Renner, Entropy accumulation, *Commun. Math. Phys.* **379**, 867 (2020).
 - [11] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Phys. Phys. Fiz.* **1**, 195 (1964).

- [12] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
- [13] B. S. Cirel'son, Quantum generalizations of Bell's inequality, *Lett. Math. Phys.* **4**, 93 (1980).
- [14] P. Mironowicz and M. Pawłowski, Robustness of quantum-randomness expansion protocols in the presence of noise, *Phys. Rev. A—At., Mol. Opt. Phys.* **88**, 032319 (2013).
- [15] L. Woollerton, P. Brown, and R. Colbeck, Tight analytic bound on the trade-off between device-independent randomness and nonlocality, *Phys. Rev. Lett.* **129**, 150403 (2022).
- [16] 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 (1969).
- [17] A. J.-M. Seguinard, A. Piveteau, P. Mironowicz, and M. Bourennane, Experimental certification of more than one bit of quantum randomness in the two inputs and two outputs scenario, *New J. Phys.* **25**, 113022 (2023).
- [18] F. Dupuis and O. Fawzi, Entropy accumulation with improved second-order term, *IEEE Trans. Inf. Theory* **65**, 7596 (2019).
- [19] P. J. Brown, S. Ragy, and R. Colbeck, A framework for quantum-secure device-independent randomness expansion, *IEEE Trans. Inf. Theory* **66**, 2964 (2019).
- [20] P. Mironowicz and M. Bourennane, Finite-size security analysis for quantum protocol: A python framework using the entropy accumulation theorem with graphical interface, *ArXiv:2506.18888*.
- [21] M. Navascués, S. Pironio, and A. Acín, Bounding the set of quantum correlations, *Phys. Rev. Lett.* **98**, 010401 (2007).
- [22] M. Navascués, S. Pironio, and A. Acín, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, *New J. Phys.* **10**, 073013 (2008).
- [23] P. Mironowicz, Semi-definite programming and quantum information, *J. Phys. A: Math. Theor.* **57**, 163002 (2024).
- [24] P. Mironowicz and M. Bourennane, When more light means less quantum: Modeling Bell inequality degradation from accidental counts, *ArXiv:2507.20596*.
- [25] C. A. Miller and Y. Shi, Universal security for randomness expansion from the spot-checking protocol, *SIAM J. Comput.* **46**, 1304 (2017).
- [26] B. Chor and O. Goldreich, Unbiased bits from sources of weak randomness and probabilistic communication complexity, *SIAM J. Comput.* **17**, 230 (1988).
- [27] J. Håstad, R. Impagliazzo, L. Levin, and M. Luby, A pseudorandom generator from any one-way function, *SIAM J. Comput.* **28**, 1364 (1999).
- [28] R. König, R. Renner, and C. Schaffner, The operational meaning of min- and max-entropy, *IEEE Trans. Inf. Theory* **55**, 4337 (2009).
- [29] I. Issa and A. B. Wagner, Measuring secrecy by the probability of a successful guess, *IEEE Trans. Inf. Theory* **63**, 3783 (2017).
- [30] P. Brown, H. Fawzi, and O. Fawzi, Device-independent lower bounds on the conditional von Neumann entropy, *Quantum* **8**, 1445 (2024).
- [31] P. Wittek, Algorithm 950: Ncpol2sdpa—sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables, *ACM Trans. Math. Softw. (TOMS)* **41**, 1 (2015).
- [32] MOSEK Modeling Cookbook. 3.2.3 ed., 2021.
- [33] L. Trevisan, Extractors and pseudorandom generators, *J. ACM* **48**, 860 (2001).
- [34] A. De, C. Portmann, T. Vidick, and R. Renner, Trevisan's extractor in the presence of quantum side information, *SIAM J. Comput.* **41**, 915 (2012).
- [35] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, Leftover hashing against quantum side information, *IEEE Trans. Inf. Theory* **57**, 5524 (2011).
- [36] L. K. Shalm, Y. Zhang, J. C. Bienfang, C. Schlager, M. J. Stevens, M. D. Mazurek, C. Abellán, W. Amaya, M. W. Mitchell, M. A. Alhejji, *et al.*, Device-independent randomness expansion with entangled photons, *Nat. Phys.* **17**, 452 (2021).
- [37] W. Zhang, T. van Leent, K. Redeker, R. Garthoff, R. Schwonnek, F. Fertig, S. Eppelt, W. Rosenfeld, V. Scarani, C. C.-W. Lim, *et al.*, A device-independent quantum key distribution system for distant users, *Nature* **607**, 687 (2022).
- [38] D. P. Nadlinger, P. Drmota, B. C. Nichol, G. Araneda, D. Main, R. Srinivas, D. M. Lucas, C. J. Ballance, K. Ivanov, E.-Z. Tan, *et al.*, Experimental quantum key distribution certified by Bell's theorem, *Nature* **607**, 682 (2022).
- [39] R. Bhavsar, S. Ragy, and R. Colbeck, Improved device-independent randomness expansion rates using two sided randomness, *New J. Phys.* **25**, 093035 (2023).