More Randomness from the Same Data

Jean-Daniel Bancal,[1] Lana Sheridan,[1] and Valerio Scarani[1, 2]

[1]*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117543*
[2]*Department of Physics, National University of Singapore, 3 Science Drive 2, Singapore 117542*

Correlations that cannot be reproduced with local variables certify the generation of private randomness. Usually, the violation of a Bell inequality is used to quantify the amount of randomness produced. Here, we show how private randomness generated during a Bell test can be directly quantified from the observed correlations, without the need to process these data into an inequality. The frequency with which the different measurement settings are used during the Bell test can also be taken into account. This improved analysis turns out to be very relevant for Bell tests performed with a finite collection efficiency. In particular, applying our technique to the data of a recent experiment [Christensen *et al.*, Phys. Rev. Lett. 111, 130406 (2013)], we show that about twice as much randomness as previously reported can be potentially extracted from this setup.

I. RANDOMNESS FROM BELL TESTS

A. Introduction

Sources of true randomness have numerous applications, be they in algorithms, gambling, sampling, or cryptography [1]. However, the unpredictability of these sources is typically difficult to certify since they may be correlated to external variables, and access to those other variables could render them predictable. Therefore, there is strong motivation for developing sources of randomness that can be certified as being uncorrelated to any outside process or variable, i.e. sources of private randomness. Quantum physics offers this opportunity when a Bell inequality is violated. Since this link between nonlocal correlations and randomness was highlighted [2], it was used for randomness generation or expansion [3–6] and randomness amplification [7–9].

All previous works certify randomness through the amount of violation of one or more Bell inequalities [10]. As such, they only take into account a coarse grained description of the knowledge available in a device-independent experiment. Moreover, in the case of randomness generation or expansion, they quantified the amount of randomness by analysing each measurement setting separately (or by computing lower bounds that assume a biased choice of measurement settings). Here, we propose a method to certify rates of randomness generation in the limit of large aquisitions directly from the full expected statistics, i.e. both the outcomes correlations and the frequency with which measurement settings are chosen. We show that this tool provides immediate improvements on the quantification of private randomness by applying it in the context of Bell tests with finite detection efficiency.

B. Working assumptions

Randomness can be studied under different sets of assumptions. The situation we have in mind here is the generic case of an experimental physicist who wants to convince a colleague that his setup is doing its job properly. We thus choose our working assumptions accordingly.

Namely, we consider in this paper that the devices used to produce randomness were acquired from a trusted provider (or built by the trustable experimental physicist himself), meaning that the devices may not operate as hoped for, but they are precluded from being actively malicious: for instance, they don't incorporate an RF transmitter that could leak supposedly private information to the outside world. There is also no adversary Eve in this scenario, but only a skeptic verifier Thomas (the colleague), who wants unquestionable evidence that genuine private randomness is being produced. Thomas does not hold quantum side information about the devices (the honest provider didn't purposely design the devices so that they would remain entangled to a third system); but he may hold a more detailed classical description of the devices than that available to average users and to the provider. If this classical information is sufficient to predict the outcomes produced by the devices, because they act in a deterministic way for instance, then Thomas will conclude that there is no intrinsic randomness. In particular, pseudo-random sources would be detected in this way, since they are totally predictable given their classical description. This *trusted-provider assumption* was first explicitly stated in [4], and has been used to assess recent experiments on randomness expansion [3, 11].

An important concern in randomness analysis is *measurement independence*, i.e. that measurement settings chosen in each run are independent of the state of the quantum systems measured in that run. Measurement independence can only be guaranteed with some element of trust; it has been assumed in the first studies, then relaxed in several works [7–9, 12–15]. In our trusted-provider scenario, full measurement independence holds as a natural consequence, even when the settings are chosen with a public pseudorandom string (and can thus be known by the verifier), because one trusts that the provider didn't build the devices with any particular string of measurement choices in mind. In this context one can thus talk of private randomness *genera-*

tion, because the randomness certified here can be generated without using any initial private randomness. This contrasts with device-independent randomness expansion (DIRE) schemes, which rely on the consumption of initial private randomness [4].

Finally, for simplicity, we also assume in this paper that all measurements are performed a large enough number of times so that the corrections of finite statistics can be safely neglected. In this asymptotic limit and under measurement independence, memory effects do not matter if the observed statistics behave like a supermartingale [16]. Assuming that this will be the case in a practical realization, we can formally restrict ourselves to the case where subsequent runs of the experiment produce independent and identically distributed (i.i.d.) variables. Finite statistics corrections, even accounting for memory effects, could be implemented on top of these results as demonstrated in [3, 16, 17].

II. RANDOMNESS FROM OBSERVED STATISTICS

It is common, and convenient, to estimate the amount of private randomness that can be extracted from some correlations in terms of a Bell inequality violation [3]. Indeed, violation of a Bell inequality certifies that no pre-established strategy can reproduce the observed correlations, and thus, that no one can perfectly predict them. However, a Bell inequality violation only constitutes a partial knowledge of all the statistics produced during a Bell type experiment. Since in general the full correlation statistics can be known, we ask here how much randomness can be certified based on this more complete knowledge.

A. Full observed correlations

Just as quantum states can exhibit two kinds of randomness, the intrinsic randomness of pure states, and the randomness of mixed states [18], correlations can also exhibit these two kinds of randomness [19], the intrinsic randomness being uncorrelated to outside systems, and thus private. For instance, any *extremal* quantum correlations $P(a, b|x, y)$, which cannot be decomposed as a convex mixture of other correlations, hold intrinsic randomness: the maximum probability with which an external observer is able to guess the outcomes observed by the parties when they perform measurements (x, y) is indeed given by the probabilities $P(ab|xy)$ themselves through:

$$G_{x,y} = \max_{a,b} P(ab|xy). \qquad [P \text{ extremal}] \qquad (1)$$

The associated randomness is then $H_{\min}(AB|XY) = -\log_2(G_{xy})$. Here a, b denote outcomes that two parties,

Alice and Bob can observe when they perform measurements x, y on their respective systems.

Conversely, when some correlations $P(ab|xy)$ can be decomposed as a convex mixture of other quantum correlations, one cannot a priori exclude that part of the outcomes' indeterminacy is simply due to classical lack of knowledge of Alice and Bob. In the trusted-provider assumptions mentioned earlier, which ensures that the provider of the boxes used by Alice and Bob keeps no state entangled with the devices in his factory, the quantum state shared by the parties and the verifier Thomas takes the form of a quantum-classical (q-c) state:

$$\rho_{ABT} = \sum_\lambda q_\lambda \rho_{AB}^{(\lambda)} \otimes |\lambda\rangle \langle\lambda| . \qquad (2)$$

Here λ is a classical variable potentially known to the verifier which can help him refine his guessing strategy. It provides the most detailed information when the $\rho_{AB}^{(\lambda)}$ are pure states. In particular, the verifier does not hold a purification of ρ_{AB}, all his side-information is classical. Moreover, as discussed above, ρ_{ABT} can be assumed to be identical for each run.

Given the form of the state ρ_{ABT}, the correlation observed by the parties can be written as

$$P(ab|xy) = \sum_\lambda q_\lambda P_\lambda(ab|xy). \qquad (3)$$

For an arbitrary set of correlations (i.e. possibly not extremal ones) the average probability with which Thomas can guess the outcomes observed by Alice and Bob correctly, given his knowledge of the inputs x, y and of λ, is thus given by the maximum of

$$G_{x,y}(P|\{q_\lambda, P_\lambda\}_\lambda) = \sum_\lambda q_\lambda \max_{a,b} P_\lambda(ab|xy) \qquad (4)$$

over all convex decompositions (3) that reproduce, on average, the observed correlations $P(ab|xy)$.

At first, it is not clear whether such an expression could be optimized easily, because it might involve an arbitrary number of λ's. The following proposition is useful in this respect (proof in Appendix A).

Proposition 1. Without loss of generality, the maximum value of

$$\sum_\lambda q_\lambda \max_{a,b} \sum_{x,y,\mu} f(x,y,\mu) P_\lambda(a(x,y,\mu), b(x,y,\mu)|xy)$$
$$(5)$$

over all decompositions of the form (3) can be achieved by considering one term in this decomposition per possible argument of the inner maximization. Here μ is a discrete variable, and $f(x,y,\mu)$, $a(x,y,\mu)$, $b(x,y,\mu)$ are functions defined on the support of $f(x,y,\mu)$.

To clarify, the arguments of the maximization are each value that the outcomes can take, given each value the inputs x, y, and perhaps some other variable μ, can take. So, in the specific case of evaluating equation (4), the

support of $f(x, y, \mu)$ consists of one point, and the summation over x, y, and μ is effectively already subsumed into P_λ. Then if the number of possible outcomes for Alice's measurements is given by $|a|$, and similarly $|b|$ denotes Bob's number of possible outcomes, this proposition ensures that no more than $|a| \cdot |b|$ terms need to be considered in the decomposition (3) when optimizing (4). Setting $\lambda = (\alpha, \beta)$, with $\alpha = \{0 \ldots |a| - 1\}$, $\beta = \{0 \ldots |b| - 1\}$ one can thus use the following program to compute the amount of private randomness that can be extracted from the correlations $P(ab|xy)$ when settings x, y are used:

$$G_{x,y}(P) = \max_{P_{\alpha\beta}} \sum_{\alpha\beta} P_{\alpha\beta}(\alpha\beta|xy)$$

$$\text{s.t.} \sum_{\alpha\beta} P_{\alpha\beta}(ab|xy) = P(ab|xy) \qquad (6)$$

$$P_{\alpha\beta}(ab|xy) \text{ is quantum,}$$

where the sum over α, β indicates that there is one value of λ for each value of α and β, and we have absorbed the weights $q_{\alpha\beta}$ into the normalization of the probability distributions $P_{\alpha\beta}$.

Notice that this expression uses the observed distribution $P(ab|xy)$ directly rather than the value of a Bell inequality.

B. Taking the input distribution into account

So far we considered the randomness that can be certified in presence of correlations $P(ab|xy)$ when a given set of settings (x, y) are used. This analysis can be straightforwardly generalized to the situation in which the inputs are chosen with some arbitrary probability $p(x, y)$ by noting that the average guessing probability in this case reads

$$G(P|\{q_\lambda, P_\lambda\}_\lambda) = \sum_{xy} p(x, y) G_{x,y}(P|\{q_\lambda, P_\lambda\}_\lambda)$$

$$= \sum_\lambda q_\lambda \max_{\mathbf{a},\mathbf{b}} \sum_{xy} p(x, y) P_\lambda(\mathbf{a}_{xy}\mathbf{b}_{xy}|x, y)$$

$$(7)$$

where $\mathbf{a}_{xy} \in \{0, \ldots, |a| - 1\}$ and $\mathbf{b}_{xy} \in \{0, \ldots, |b| - 1\}$ for all $x = \{0, \ldots, |x| - 1\}$ and $y = \{0, \ldots, |y| - 1\}$ denote the outcomes considered for each of the settings x, y.

Thanks to proposition 1 it is sufficient to consider only at most $(|a||b|)^{|x||y|}$ λ's when optimizing (7) over all decompositions of the form (3) (since the support of $f(x, y, \mu) = p(xy)$ in proposition 1 is $|x||y|$ distinct points). The amount of private randomness that can be certified when some correlations are observed by choosing settings according to the distribution $p(x, y)$ can thus be quantified by the following program:

$$G(p, P) = \max_{P_{\alpha\beta}} \sum_{xy} p(x, y) \sum_{\alpha\beta} P_{\alpha\beta}(\alpha_{xy}\beta_{xy}|xy)$$

$$\text{s.t.} \sum_{\alpha\beta} P_{\alpha\beta}(ab|xy) = P(ab|xy) \qquad (8)$$

$$P_{\alpha\beta}(ab|xy) \text{ is quantum,}$$

where the weights are again absorbed into the normalizations of $P_{\alpha\beta}$ and we have $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_{00}, \ldots \boldsymbol{\alpha}_{|x|-1,|y|-1}\}$ with $\boldsymbol{\alpha}_{xy} \in \{0, \ldots, |a| - 1\}$ and likewise for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{xy}$, i.e. bold greek letters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote vectors taking one of $|a|^{|x||y|}$ or $|b|^{|x||y|}$ possible values (as indexed by the components $\boldsymbol{\alpha}_{xy}$ and $\boldsymbol{\beta}_{xy}$).

The idea behind this program is that Thomas could know the relative frequencies of the inputs and happen to hold a description of the best decomposition for this case. However he does not have active access to the boxes: under measurement independence, the decomposition cannot depend on the particular measurement that is performed in each run, and so the decomposition (3) must be identical for all choices of settings x, y.

C. Getting practical bounds and certificate

In order to perform the optimizations (8) (of which (6) is a special case) one needs to describe accurately the set of quantum correlations. Since no method is known for this, we upperbound the guessing probability G by relaxing the constraint that $P_{\alpha\beta}(ab|xy)$ be quantum, to let it belong to some level of the NPA hierarchy [20, 21]. This turns equations (6) and (8) into semidefinite programs which can be efficiently computed numerically, with the guarantee of approaching the exact quantum value as the level of hierarchy increases.

Alternatively, one could also replace this condition by requiring only that the correlations $P_{\alpha\beta}(ab|xy)$ be non-signalling. In this case the optimization takes the form of a single linear program, and describes the randomness exatractable from correlations in a world described by a generalized no-signalling theory rather than quantum theory.

Apart from computing the amount of randomness that can be extracted with respect to a verifier in different contexts, one can show that these programs also provide, through their dual, the description of a certificate for this conclusion. This certificate takes the form of a Bell expression, whose expectation value on the tested correlations $P(ab|xy)$ can only be observed in presence of the amount of randomness found by the program itself. To see this (full details can be found in Appendix B), we note that the optimization (8) can be straightforwardly adapted to estimate the amount of randomness that can be certified from a Bell inequality violation. Namely, if $I = \sum_{abxy} c_{abxy} P(ab|xy)$ is a Bell expression then the maximum guessing probability achievable by a verifier

when the Bell value v is observed is given by:

$$G(v) = \max_{P_{\alpha\beta}} \sum_{xy} p(x,y) \sum_{\alpha\beta} P_{\alpha\beta}(\boldsymbol{\alpha}_{xy}\boldsymbol{\beta}_{xy}|xy)$$
$$\text{s.t.} \sum_{abxy\alpha\beta} c_{abxy} P_{\alpha\beta}(ab|xy) = v \tag{9}$$
$$\sum_{ab\alpha\beta} P_{\alpha\beta}(ab|xy) = 1$$
$$P_{\alpha\beta}(ab|xy) \text{ is quantum.}$$

Note that a similar program using only one value of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ was already presented in [3] to evaluate the randomness associated to a Bell inequality violation with a biased choice of inputs. However, its application relies on the concavity of the obtained function (or on its concave hull) [18], which is not always guaranteed (see Appendix C for an example of non-concave function obtained in this way). The application of that program thus requires in general knowledge of the function for all violation v. Here, program (9) takes care of the concavity requirements through the use of several λ's. It thus directly produces the guessing probability relevant for randomness estimation.

III. RESULTS

A. Role of the inputs distribution

To illustrate the role of the inputs distribution in randomness certification, we compare in figure 1 the randomness that can be certified in presence of a CHSH violation through (9) when one pair of settings is preferentially chosen, or when all settings are chosen with equal probability [31]. Significantly more randomness (about twice as much) can be certified in presence of nonmaximal CHSH violation if the settings are chosen uniformly.

This can be understood by considering that when the randomness is extracted from a fixed choice of settings, one cannot exclude that the decomposition (3) is most adapted for the verifier to guess the outcomes observed when those settings are precisely used. However such a decomposition need not be optimal to guess the outcomes observed when performing other measurements. Since the decomposition is the same independently of which settings are used, it follows that the average guessing probability over the different settings is reduced.

B. Application to finite efficiency Bell tests

In practice, certifying private randomness generation in a Bell experiment requires a reasonable separation between the measured subsystems, as well as closure of the detection loophole. Without these conditions fulfilled, the possibility remains for the verifier to find a way of
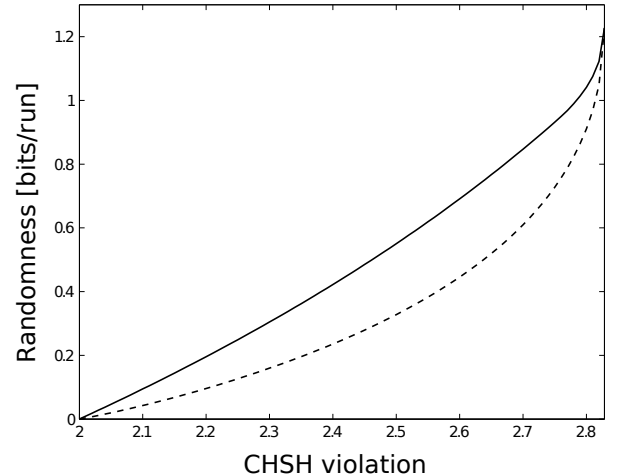

FIG. 1: Private randomness that can be certified as a function of CHSH violation in the fully biased input case (dashed line) and in the uniform choice of inputs case (full line).
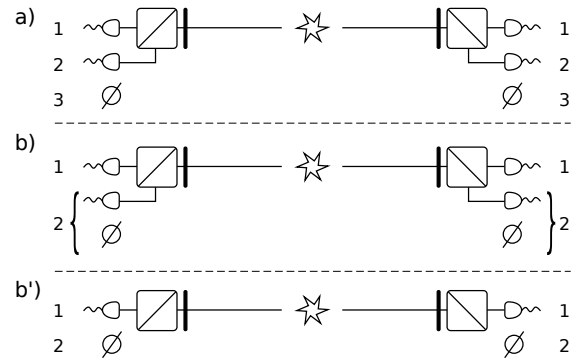

FIG. 2: Lossy Bell tests with different numbers of detectors and possible outcomes. a) With two detectors on each side, it is possible to attribute a distinct outcome to the result of each successful measurement, while identifying no-detection events. This is the most general treatment of lossy Bell test, but it requires the usage of three outcomes on each side. b) In order to recover two outcomes only, the no-detection events must be associated with one of the detector. b') The same statistics as produced in case b) can be obtained by using only one detector on each side. In this case, a click triggers the first outcome while the second outcome is produced in case of no-detection.

guessing the outcomes observed during the experiment, which then cannot be guaranteed private [22].

These constraints on parties separation and on their detection efficiency are still very demanding experimentally today. With photonic systems for instance, overall detection efficiencies which are sufficient to close the detection loophole have just been recently demonstrated [11, 23]. It is thus experimentally relevant to ask what is the best way to certify private randomness in presence of finite efficiency detectors.
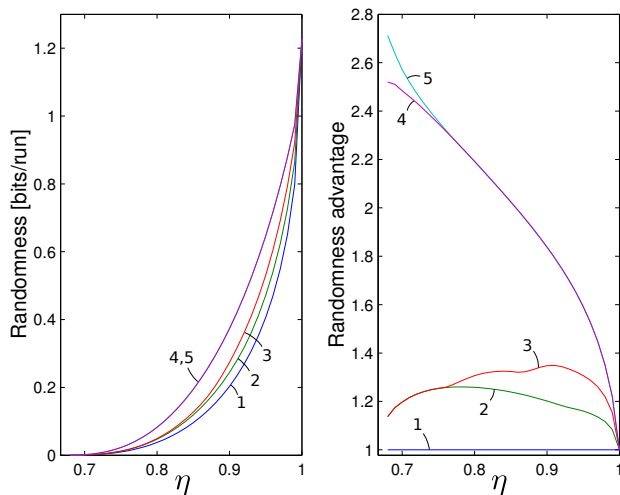
FIG. 3: Comparison of the randomness that can be certified in an Eberhard experiment under different circumstances. Curves 1,2,3 show the bound obtained when extracting randomness from a specific choice of settings, whereas curves 4 and 5 assume a uniform choice of settings. Curves 1 and 4 based on the observed CHSH violation were computed with (9) (curve 1 serves as a reference with a normalization of 1 in the second figure). Curves 2 and 5 take into account the full 2-outcomes statistics and were computed with (6) and (8). For $\eta \lesssim 90\%$ the amount of randomness displayed by curve 2 can be certified with inequality (10). Curve 3, based on the full 3-outcomes statistics was computed with (8). Bounds were computed with the Sedumi solver [25] at local level 1 of the NPA hierarchy for uniform choice of settings, and with a higher level for based inputs [20, 26].

In the simplest Bell experiment in which two parties can each use two binary measurements, finite detection efficiency can be dealt with in several ways (see Fig. 2). Either a third outcome is introduced for both Alice and Bob and all photons that go undetected are assigned to the third outcome, or all undetected photons are assigned to one of the two existing outcomes. The first solution is more general, but requires the usage of three-outcome Bell inequalities and two efficient detectors each for Alice and Bob, while the second solution can be treated in the two-outcome paradigm and has the advantage of only requiring one detector on each side. Here we consider the two cases, with all detectors having the same efficiency $\eta = \eta_A = \eta_B$.

Eberhard [24] showed that whenever $\eta > 2/3$, it is possible to violate the CHSH inequality by measuring partially-entangled states of the form $|\psi\rangle = \cos\theta |00\rangle + \sin\theta |11\rangle$ with settings of the form $A_0 = \cos\alpha_1\sigma_z - \sin\alpha_1\sigma_x$, $A_1 = \cos\alpha_2\sigma_z + \sin\alpha_2\sigma_x$, $B_0 = \cos\alpha_1\sigma_z + \sin\alpha_1\sigma_x$, $B_1 = \cos\alpha_2\sigma_z - \sin\alpha_2\sigma_x$ where θ, α_1, α_2 are functions of η. Here we analyse how much randomness can be extracted in presence of these correlations. The results are summarized in figure 3.

As shown in the figure, different amounts of randomness can be certified depending of the approach used.

The optimal bounds are obtained when choosing measurement settings uniformly. In this case the CHSH inequality is the optimal randomness witness whenever the efficiency is $\eta \gtrsim 80\%$.

To demonstrate the improvement provided by our method in presence of finite detection efficiency, we applied it to the correlations reported in [11]. The program (8) certifies a rate of 0.00014567 bits per run, i.e. two times more than mentioned in that paper. Notice that, because of statistical fluctuations, the frequencies computed from the raw data are non-quantum and even signaling; this artefact can be corrected with a minor reformatting. (see appendix D for details).

C. A new Bell expression.

Let us now go one step back and fix a highly biased inputs distribution (i.e. let's consider extracting randomness from a fixed set of inputs). When $\eta < 1$, the CHSH inequality does not certify the largest amount of randomness. The most randomness is obtained by considering the three-outcomes correlations, but already the bounds computed from the two-outcomes correlations show an improvement.

As it turns out, for efficiencies smaller than $\sim 90\%$, we can provide a Bell inequality that is optimal, in the sense that its amount of violation certifies the optimal amount of randomness. This inequality is found using the dual of (6) to be

$$
\begin{aligned}
V &= |\gamma\langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \langle A_1 B_1\rangle| \\
&\leq \max(1+\gamma, 3-\gamma) \qquad (10)
\end{aligned}
$$

where the right hand side of the expression is the bound for local correlations (computed independently of the SDP dual). Note that this inequality is not a facet of the local polytope, since this scenario only has one nontrivial family of facets, given by the CHSH inequalities [27]. Yet, it is better adapted for randomness extraction in this situation than the CHSH inequality.

In appendix E we give a brief description of this inequality, and in particular give an analytical estimate of the intrinsic randomness that can be certified on the marginal correlations for this inequality. This analysis reveals that available analytical techniques useful to certify randomness in the marginal outcomes [3], are not adapted to this inequality. The reason for this is that they treat all inputs identically, but a symmetry between the inputs is broken in inequality (10) whenever $\gamma \neq 1$. In particular, they assume that measurements can be chosen along σ_z, whereas maximal violation of this inequality with non-maximally entangled state of the form $\cos\theta |00\rangle + \sin\theta |11\rangle$ can require measurement settings not aligned with the z axis of the Bloch sphere.

IV. CONCLUSIONS

Private randomness is a valuable resource. We have presented here a way to bound the private randomness generated during a Bell test by directly using knowledge of the full outcomes correlations and of the inputs' choice distribution. Our analysis is concerned with large data sets and makes some arguably reasonable assumptions about the devices used in the Bell experiment, but does not require a characterization of these devices. Working directly from the observed correlations allows us to find tighter bounds on the min-entropy of the output than previously known, thus certifying more randomness from identical experiments. We demonstrate this explicitly for the results of [11], effectively doubling the expected private randomness production rate.

This approach also furnishes new Bell inequalities through semi-definite duality to bound the guessing probability. These inequalities may not be facets of the local polytope (inequality (10) is not), but they are the ones suited to randomness extraction for the form of the correlations supplied.

Even though we focused here on bipartite Bell experiments, our result generalizes directly to scenarios involving arbitrary alphabet size of inputs and outputs and any number of parties. The tools presented here can also be applied under different sets of assumptions. For instance, they can provide bounds in presence of an untrusted provider keeping a purification of the boxes if the inputs are made public only after the purification decohered.

It would be relevant to extend the bounds presented here to the case of finite statistics, including devices with memory, possibly using methods similar to [3, 16, 17]. This work also shows that new ways of obtaining analytical bounds on the randomness need to be developped to certify optimal rates of randomness generation analytically.

Note added. While writing this article we became aware of a similar work by O. Nieto-Silleras, S. Pironio and J. Silman [28].

Appendix A: Proof of propositions 1

Let us show that whenever the best choice of the function $a(x,y,\mu)$ and $b(x,y,\mu)$ for the inner optimization of (5) is identical for two strategies λ' and λ'', i.e.

$$
\begin{aligned}
&\underset{a,b}{\mathrm{argmax}} \sum_{x,y,\mu} f(x,y,\mu) P_{\lambda'}(a(x,y,\mu),b(x,y,\mu)|xy) \\
&= \underset{a,b}{\mathrm{argmax}} \sum_{x,y,\mu} f(x,y,\mu) P_{\lambda''}(a(x,y,\mu),b(x,y,\mu)|xy),
\end{aligned}
\tag{A1}
$$

then the two stategies can be grouped together. This implies that it is sufficient to consider one λ for each possible argument of this maximum.

Let us thus assume that two strategies λ', λ'' in the considered convex decomposition of $P(ab|xy)$ are such that $\max_{a,b} \sum_{x,y,\mu} f(x,y,\mu) P_{\lambda'}(a(x,y,\mu),b(x,y,\mu)|xy)$ is achieved for the functions $a'(x,y,\mu), b'(x,y,\mu)$ and $\max_{a,b} \sum_{x,y,\mu} f(x,y,\mu) P_{\lambda''}(a(x,y,\mu),b(x,y,\mu)|xy)$ is achieved for the functions $a''(x,y,\mu) = a'(x,y,\mu)$, $b''(x,y,\mu) = b'(x,y,\mu)$. We now define a new decomposition for $P(ab|xy) = \sum_\lambda \tilde{q}_\lambda \tilde{P}_\lambda(ab|xy)$ by choosing

$$
\tilde{q}_\lambda = \begin{cases} q_\lambda & \text{if } \lambda \neq \lambda', \lambda'' \\ q_{\lambda'} + q_{\lambda''} & \text{if } \lambda = \lambda' \\ 0 & \text{if } \lambda = \lambda'' \end{cases}
\tag{A2}
$$

and

$$
\tilde{P}_\lambda(ab|xy) = \begin{cases} P_\lambda(ab|xy) & \text{if } \lambda \neq \lambda', \lambda'' \\ \frac{q_{\lambda'} P_{\lambda'}(ab|xy) + q_{\lambda''} P_{\lambda''}(ab|xy)}{q_{\lambda'} + q_{\lambda''}} & \text{if } \lambda = \lambda' \\ 0 & \text{if } \lambda = \lambda''. \end{cases}
\tag{A3}
$$

Clearly, the new decomposition is a valid convex decomposition since it satisfies $\tilde{q}_\lambda \geq 0$, $\sum_\lambda \tilde{q}_\lambda = 1$, and $P(ab|xy) =$

$\sum_\lambda \tilde{q}_\lambda \tilde{P}_\lambda(ab|xy)$. We are thus just left to show that it provides the same value for Eq. (5). This is verified through:

$$
\sum_\lambda \tilde{q}_\lambda \max_{a,b} \sum_{x,y,\mu} f(x,y,\mu)\tilde{P}_\lambda(a(x,y,\mu),b(x,y,\mu)|xy) = \sum_{\lambda\neq\lambda',\lambda''} q_\lambda \max_{a,b} \sum_{x,y,\mu} f(x,y,\mu)P_\lambda(a(x,y,\mu),b(x,y,\mu)|xy)
$$
$$
+ \tilde{q}_{\lambda'} \max_{a,b} \sum_{x,y,\mu} f(x,y,\mu)\tilde{P}_{\lambda'}(a(x,y,\mu),b(x,y,\mu)|xy)
$$
$$
= \sum_\lambda q_\lambda \max_{a,b} \sum_{x,y,\mu} f(x,y,\mu)P_\lambda(a(x,y,\mu),b(x,y,\mu)|xy).
$$

(A4)

Thus, it suffices to consider only as many strategies λ as the number of arguments to the maximization. $\square$

Appendix B: A Bell expression to certify optimal randomness extraction

Here we show that whenever the primal and dual objective functions for the dual of the SDP relaxations of Eq. (8) (of which (6) is a special case) coincide, variables of the dual provide the description of a Bell expression which can be used to certify that no one can guess the outcomes observed by the parties with a probability higher than that given by the relaxation's primal.

For this, let us write the SDP relaxation of Eq. (8), as well as its dual program [29]. For any hierarchy level, this can be done by considering for each probability distribution $P_{\alpha\beta}$ a matrix of the form $\Gamma_{\alpha\beta} = \sum_i F_i \mathrm{x}^i_{\alpha\beta}$, where F_i are constant matrices and $\mathrm{x}^i_{\alpha\beta}$ are a finite number of variables [20, 21]. For convenience, we also introduce the constant matrix $F(ab|xy)$ to pick up the terms in the Γ matrices associated to the probabilities, i.e. $P_{\alpha\beta}(ab|xy) = \mathrm{Tr}(F(ab|xy)\Gamma_{\alpha\beta})$, as well as the indicative function $f_i(ab|xy) = \mathrm{Tr}(F(ab|xy)F_i)$. The relaxation and its dual then read:

Primal:
$$
G(P) \leq \max_{\mathrm{x}^i_{\alpha\beta}} \sum_{xy} p(x,y)\sum_{\alpha\beta i} f_i(\boldsymbol{\alpha\beta}|xy)\mathrm{x}^i_{\alpha\beta}
$$
$$
\text{s.t.} \sum_{\alpha\beta i} f_i(ab|xy)\mathrm{x}^i_{\alpha\beta} = P(ab|xy)
$$
$$
\sum_i F_i \mathrm{x}^i_{\alpha\beta} \geq 0,
$$
(B1)

Dual:
$$
\min_{c_{abxy},M_{\alpha\beta}} \sum_{abxy} c_{abxy}P(ab|xy)
$$
$$
\text{s.t.} \sum_{abxy} f_i(ab|xy)c_{abxy} + \mathrm{Tr}(F_i M_{\alpha\beta})
$$
$$
= \sum_{xy} p(x,y)f_i(\boldsymbol{\alpha\beta}|xy)
$$
$$
M_{\alpha\beta} \leq 0.
$$
(B2)

Eq. (B2) is the dual of Eq. (B1) in the sense that the value of its objective function for any set of variables $(c_{abxy}, M_{\alpha\beta})$ satisfying its constraints sets an upper bound on the optimal value of the primal's optimization. In practice, solving these programs typically yields identical values for both objective functions, thus certifying their optimality. This was the case for all programs solved for this paper.

Let us now show that when the two objective functions coincide the coefficients c_{abxy} define a Bell expression, whose value achieved by the tested probabilities $P(ab|xy)$, can only be achieved when the guessing probability is bounded by the optimum of the above primal. For this, we also write an SDP relaxation of equation (9). Bounding the maximum solution of this program for any quantum correlations achieving the value v for the Bell expression defined by c_{abxy} will conclude the proof.

This relaxation is done as above by associating to each probability distribution $P_{\alpha\beta}$ a matrix $\Gamma'_{\alpha\beta} = \sum_i F_i \mathrm{y}^i_{\alpha\beta}$. The obtained semi-definite program then reads:

$$
G(v) \leq \max_{\mathrm{y}^i_{\alpha\beta}} \sum_{xy} p(x,y)\sum_{\alpha\beta i} f_i(\boldsymbol{\alpha\beta}|xy)\mathrm{y}^i_{\alpha\beta}
$$
$$
\text{s.t.} \sum_{abxy}\sum_{\alpha\beta i} c_{abxy} f_i(ab|xy)\mathrm{y}^i_{\alpha\beta} = v
$$
$$
\sum_{ab\alpha\beta i} f_i(ab|xy)\mathrm{y}^i_{\alpha\beta} = 1
$$
$$
\sum_i F_i \mathrm{y}^i_{\alpha\beta} \geq 0.
$$
(B3)

So let us bound $G(v)$ for the coefficients c_{abxy} of the above dual, by using the other variables of this dual at optimum. For this we write

$$
\begin{aligned}
G(v) &\leq \max_{\mathrm{y}^i_{\boldsymbol{\alpha\beta} i}} \sum_{xy} p(x,y) \sum_{\boldsymbol{\alpha\beta}} f_i(\boldsymbol{\alpha\beta}|xy)\mathrm{y}^i_{\boldsymbol{\alpha\beta}} \\
&= \max_{\mathrm{y}^i_{\boldsymbol{\alpha\beta} i}} \sum_{\boldsymbol{\alpha\beta}} \left[\sum_{abxy} f_i(ab|xy)c_{abxy} + \mathrm{Tr}(F_i M_{\boldsymbol{\alpha\beta}}) \right] \mathrm{y}^i_{\boldsymbol{\alpha\beta}} \\
&= \max_{\mathrm{y}^i_{\boldsymbol{\alpha\beta}}} v + \sum_{\boldsymbol{\alpha\beta}} \mathrm{Tr}\left(\left(\sum_i F_i \mathrm{y}^i_{\boldsymbol{\alpha\beta}} \right) M_{\boldsymbol{\alpha\beta}} \right) \\
&\leq v = \max_{\mathrm{x}^i_{\boldsymbol{\alpha\beta}}} \sum_{xy} p(x,y) \sum_{\boldsymbol{\alpha\beta} i} f_i(ab|xy)\mathrm{x}^i_{\boldsymbol{\alpha\beta}},
\end{aligned}
\tag{B4}
$$

where we used the fact that the $M_{\boldsymbol{\alpha\beta}}$ matrices are negative and the $\mathrm{y}^i_{\boldsymbol{\alpha\beta}}$ variables must satisfy $\sum_i F_i \mathrm{y}^i_{\boldsymbol{\alpha\beta}} \geq 0$ to get to the last line. This concludes the proof. $\qquad\square$

Appendix C: Concavity of the guessing probability function

As mentioned in the main text, it was proposed in [3] to estimate the randomness associated to a Bell inequality violation by upperbounding the guessing probability $G_{xy}(v)$ by the concave hull of the function

$$
\begin{aligned}
f_{xy}(v) = \max_{a,b,P} \quad & P(ab|xy) \\
\text{s.t.} \quad & \sum_{abxy} c_{abxy} P(ab|xy) = v \\
& P(ab|xy) \text{ is quantum.}
\end{aligned}
\tag{C1}
$$

When relaxing in this optimization the quantum condition to some level of the NPA hierarchy, computing the function $f_{xy}(v)$ amount to solving $|a| \cdot |b|$ semidefinite programs involving a matrix of size N (one optimization for each value of (a,b)), and taking the concave hull of the maximum of the $|a| \cdot |b|$ functions of v found in this way. Let us show that taking this concave hull is not necessarily a trivial step.

For this matter, we consider the function $f_{0,0}(v)$ for the γ inequality (10) with $\gamma = 3/4$. The bound on this function found at local level 1 of the NPA hierarchy is plotted in figure 4. Clearly it is not concave. In particular, the value found for $v = 2.4$ is $f_{00}(2.4) \simeq 0.7204$, whereas its concave hull for the same v takes the value $\simeq 0.8075$. In fact, the result of the alternative optimization (9) presented in the main text also yields this value. Since one can check that there exists a decomposition of the form (3) in terms of quantum correlations that achieve this guessing probability, it is indeed an optimal value.

Finding the correct guessing probability in this case with this method thus requires one to solve (C1) for both $v = 0$ and $v \simeq 2.607$ (analytical value unknown), which requires some global knowledge about the $f_{00}(v)$ function. In contrast, the method presented in the main text in Eq. (9) directly yields the guessing probability $G_{00}(2.4)$ by only considering the value $v = 2.4$. It requires however, for the same level of the hierarchy, one optimization with $|a| \cdot |b|$ semidefinite blocks of size N (which constitutes a larger problem than the individual ones above).

Appendix D: Application to [Christensen *et al.*, Phys. Rev. Lett. 111, 130406 (2013)]

Nonlocal correlations observed with limited detection efficiency (about 75%) are reported in [11][32]. An estimation of randomness based on these observations is then also provided, with the help of known theoretical techniques. Here we perform a similar analysis by using the method described in the main text. Namely we estimate the amount of private randomness one could expect to extract from the setup used for this experiment in the limit of many runs.

Before applying our tools to this data, we need to slightly process it in order to make it fit with our theoretical description. Indeed, being the result of a finite number of measurements, the frequencies with which the different outcomes were observed violate the no-signalling conditions, i.e.

$$
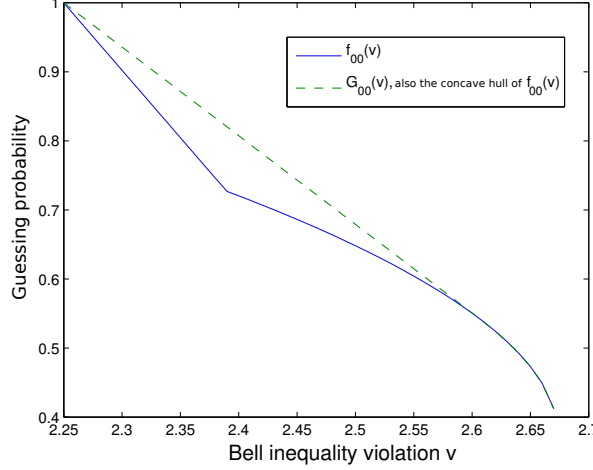\frac{S_A(a,b)}{N(a,b)} \neq \frac{S_A(a,b')}{N(a,b')}
\tag{D1}
$$

FIG. 4: Bounds on the guessing probability $G_{00}(v)$ found using the programs (C1) and (9) of the main text. Since the function $f_{00}(v)$ is not concave, its concave hull for $v = 2.4$ requires knowledge of $f_{00}(v)$ for $v = 0$ and $v \simeq 2.607$. One can check that the the function $G_{00}(v)$ coincides with the concave hull of $f_{00}(v)$.

where S_A is the number of singles observed at Alice, and N the number of events recorded in presence of some settings (we use here the notation of [11]). All quantum correlations being no-signalling, this violation prevents us from applying the optimizations presented in the main text to these frequencies directly.

Since one expects this signalling character to disappear in the limit of many measurements, we choose to remove this feature by projecting the frequencies onto the no-signalling probability subspace in an orthogonal manner. This can be done by considering the full probability space as $\mathbb{R}^{16}$ equipped with the standard inner product. Alternatively, this projection can be performed directly in terms of expectation value for the ± 1 eigenvalue observables A_{xy}, B_{xy}, as we do now: first we compute the expectation values for each setting (x, y) directly from the counts as

$$\langle A_{00} \rangle = \frac{2S_A(a, b) - N(a, b)}{N(a, b)}, \quad \langle B_{00} \rangle = \frac{2S_B(a, b) - N(a, b)}{N(a, b)}, \quad \langle A_{00}B_{00} \rangle = \frac{4C(a, b) - 2S_A(a, b) - 2S_B(a, b) + N(a, b)}{N(a, b)}$$

(D2)

and similarly for the other pairs of settings. The projection is then obtained by setting $\langle A_x \rangle = \frac{\langle A_{x0} \rangle + \langle A_{x1} \rangle}{2}$ and $\langle B_y \rangle = \frac{\langle B_{0y} \rangle + \langle B_{1y} \rangle}{2}$.

The no-signalling correlations obtained in this way take value:

$$\begin{aligned}
\langle A_0 \rangle &= -0.9966077, & \langle A_1 \rangle &= -0.9891902 \\
\langle B_0 \rangle &= -0.9965994, & \langle B_1 \rangle &= -0.9896845 \\
\langle A_{00}B_{00} \rangle &= 0.9975133, & \langle A_{01}B_{01} \rangle &= 0.9911125 \\
\langle A_{10}B_{10} \rangle &= 0.9907040, & \langle A_{11}B_{11} \rangle &= 0.9791706,
\end{aligned}$$

(D3)

yielding a CHSH violation of ~ 2.000159.

Note that the correlations (D3) could still, in principle, lie outside the set of quantum correlations, thus still preventing a direct application of our technique. We checked however that this was not the case.

In general, other projections could also be chosen to relate the observed signalling statistics to some quantum correlations. By construction, the particular projection used here minimizes the euclidean distance between the projected probabilities and the original ones. This ensures that its result cannot be too far away from the original statistics, as measured by any p-norm. In particular, it must be close to any point belonging both to the quantum set and to a confidence ball of radius ϵ in p-norm around the original probabilities.

Running the programs (8) and (9) either with the obtained correlations (D3) or the CHSH violation, and a uniform choice of settings ($p(x, y) = 1/4$) certifies in both cases a private randomness rate of 0.00014567 [bits per run] $\simeq 16207$ [bits] / 111259682 [runs]. This is about twice as much randomness per run as reported in [11][33].

Appendix E: Properties of an Inequality for Randomness Certification

Let us provide a brief description of the properties of inequality (10). For $\gamma > 1$, to increase the value that this expression takes the first correlation term should be made larger (in absolute value) at the expense of the other terms. To do this, measurements A_0 and B_0 should be chosen to have an inner product smaller than $\frac{1}{\sqrt{2}}$. This expression reaches its maximum value for the maximally entangled states $|\Phi^+\rangle$, $|\Psi^-\rangle$ and the following set of coplanar measurement angles:

$$
\begin{aligned}
\mathbf{a_0}.\mathbf{b_0} &= \cos\theta_{00} \\
\mathbf{a_0}.\mathbf{b_1} &= \cos\alpha \\
\mathbf{a_1}.\mathbf{b_0} &= \cos\alpha \\
\mathbf{a_1}.\mathbf{b_1} &= \cos(\pi - \alpha)
\end{aligned}
\tag{E1}
$$

where $\mathbf{a_i}, \mathbf{b_i}$ are coplanar vectors specifying the measurement settings A_i, B_i, and $\alpha = \frac{\pi - \theta_{00}}{3}$. The angle θ_{00} can be expressed in terms of γ as:

$$
\theta_{00}(\gamma) = 3\cos^{-1}\left(\frac{\sqrt{5 + (1/\gamma)\left(\sqrt{3}\sqrt{(3\gamma - 1)(\gamma + 1)} - 1\right)}}{2\sqrt{2}} \right).
\tag{E2}
$$

The maximum quantum value V_Q of the inequality also depends on γ:

$$
V_Q(\gamma) = \gamma\cos\left(\theta_{00}(\gamma)\right) + 3\sin\left(\frac{\pi}{6} + \frac{\theta_{00}(\gamma)}{3} \right)
\tag{E3}
$$

which is found from computing $\mathrm{Tr}(V\,|\Psi^-\rangle\langle\Psi^-|)$.

Let us now consider how to find a bound on the amount of marginal randomness that a particular value v of the Bell inequality (10) implies. Here, by marginal randomness we mean the randomness of only one party's outcome assuming that the other party's outcome is securely destroyed. Following [3, 18], we are interested in calculating the guessing probability of the most biased pure state for which $\mathrm{Tr}(V\rho) = v$. The family of biased states we require here are of the form $|\Phi(\beta)\rangle = \cos\beta\,|00\rangle + \sin\beta\,|11\rangle$.

The maximum value that this Bell inequality can take using a state of the form $|\Phi(\beta)\rangle$ is found by optimizing the measurements, A_0, A_1, B_0, B_1 in:

$$
\hat{V} = A_0(\gamma B_0 + B_1) + A_1(B_0 - B_1).
\tag{E4}
$$

The optimal violation is found when $\mathbf{a_0} = \frac{\gamma\mathbf{b_0} - \mathbf{b_1}}{|\gamma\mathbf{b_0} - \mathbf{b_1}|}$ and $\mathbf{a_1} = \frac{\mathbf{b_0} - \mathbf{b_1}}{|\mathbf{b_0} - \mathbf{b_1}|}$.

Using this observation about the optimal alignment of the measurement operators A_0 and A_1 and following the approach of [30], we can conclude that

$$
V_Q(\Phi(\beta)) = \max_{v_0, v_1} \|M\mathbf{v_0}\| + \|M\mathbf{v_1}\|
\tag{E5}
$$

where

$$
\mathbf{v_0} = \begin{pmatrix} \gamma b_x^0 + b_x^1 \\ 0 \\ \gamma b_z^0 + b_z^1 \end{pmatrix}, \text{ and } \mathbf{v_1} = \begin{pmatrix} b_x^0 - b_x^1 \\ 0 \\ b_z^0 - b_z^1 \end{pmatrix},
\tag{E6}
$$

(a symmetry allows the y values to be chosen as 0) and M is the matrix representing the state $|\Phi(\beta)\rangle$ on the Bloch sphere. M is diagonal with eigenvalues $\sin(2\beta), -\sin(2\beta), 1$. Then setting $b_x^0 = \cos\phi_0$ and $b_x^1 = \cos\phi_1$,

$$
\begin{aligned}
V_Q(\Phi(\beta)) = \max_{\phi_0, \phi_1}\Bigg[&\sqrt{(\gamma\cos\phi_0 + \cos\phi_1)^2 + \sin^2(2\beta)(\gamma\sin\phi_0 + \sin\phi_1)^2} \\
&+ \sqrt{(\cos\phi_0 - \cos\phi_1)^2 + \sin^2(2\beta)(\sin\phi_0 - \sin\phi_1)^2}\Bigg].
\end{aligned}
\tag{E7}
$$

We cannot solve this expression analytically for general γ, however for cases of practical interest, γ will be only a little larger than one.

In order to obtain an analytic estimate of the randomness, we assume that the values of γ lie in $1 \leq \gamma \lesssim 1.1$. Let us assume that within this interval the expression for the maximum violation can be expanded around $\gamma = 1$ as

$$V_Q(\Phi(\beta)) \;=\; 2\sqrt{1 + \sin^2(2\beta)} + \frac{\epsilon}{\sqrt{1 + \sin^2(2\beta)}} + \frac{\epsilon^2}{(1 + \sin^2(2\beta))^{3/2}} \left(\frac{\sin^2(2\beta)}{2} + x \right) + O(\epsilon^3), \qquad \text{(E8)}$$

where we set $\epsilon = \gamma - 1$, $x := \sin^2(2\beta)(\sin^2(2\beta)\phi_0'[1] - \phi_1'[1]) - (\phi_0'[1] - \sin^2(2\beta)\phi_1'[1])^2$ and $\phi_0'[1] := \left.\frac{\partial \phi_0}{\partial \gamma}\right|_{\gamma=1}$ and likewise for $\phi_1'[1]$. In order to upper bound $V_Q(\Phi(\beta))$, we require a bound on x. Expanding $\frac{dV_Q}{d\gamma}$ in terms of partial differentials and equating the first-order in ϵ term on both sides gives $\phi_0'[1] = \frac{1}{2}\sin^2(2\beta)(1 + 2\phi_1'[1])$. Further, since we have expanded $\cos(\phi_1)$ in the neighbourhood of $\gamma = 1$ as $\cos(\pi/2 + \epsilon\phi_1'[1])$, we should have $-\pi < \epsilon\phi_1'[1] < \pi$, since we can restrict ϕ_1 to occupy only a range over 2π. Note that this is equivalent to making the assumption that the expansion is valid for this range of γ. Then

$$\epsilon x \;\leq\; \frac{\epsilon}{(1 + \sin^2(2\beta))^{3/2}} \left(\pi \sin^4(2\beta)(1 - \sin^2(2\beta)) \;\; + \frac{\epsilon}{2} \sin^2(2\beta) \left(1 + \sin^4(2\beta) - \frac{1}{2}\sin^2(2\beta) \right) \right). \qquad \text{(E9)}$$

(Note that using the analytic bound, when $\beta = \pi/4$, $x = 1/4$, and from a numerical analysis, $x < 1.6$ for all values of β.) Then, we can supply a bound on the maximum value that the inequality (10) can take for any value of β and small values of $(\gamma - 1)$ as:

$$V_Q(\Phi(\beta)) \;\leq\; 2\sqrt{1 + \sin^2(2\beta)} + \frac{\epsilon}{\sqrt{1 + \sin^2(2\beta)}} \left(1 + \frac{\pi \sin^4(2\beta)}{1 + \sin^2(2\beta)}(1 - \sin^2(2\beta)) \right)$$

$$+ \frac{\epsilon^2 \sin^2(2\beta)}{2(1 + \sin^2(2\beta))^{3/2}} \left(1 + \sin^4(2\beta) - \frac{1}{2}\sin^2(2\beta) \right) + O(\epsilon^3), \qquad \text{(E10)}$$

Under the assumption that the expansion is valid in the considered range (this is equivalent to assuming that the derivatives of V with respect to γ are not too large), this method gives a bound on the randomness that a particular observed violation implies, and thus a quick and simple method to calculate a lower bound on the extractable intrinsic randomness of an experiment. Unfortunately, one can check that this quick bound does not reproduce the value obtained numerically using programs presented in the main text, and in particular it does not improve on the bound implied by the CHSH inequality, at least not for values of γ in the range of interest.

The reason for this is as follows. One way to bound the intrinsic min-entropy of one party's outcome, after observing a value of the inequality (10) v, is to find the largest value of β compatible with v such that $V_Q = v$ using the bound (E10). Then the guessing probability for any choice of input measurement is $G \leq \cos^2 \beta$ for $0 \leq \beta \leq \frac{\pi}{4}$. However, it is not sufficient to use that bound, because it returns a lower amount of marginal intrinsic randomness than does the same technique using instead the CHSH inequality. In actual fact, the amount of generated randomness is always *higher* than $\cos^2 \beta$ because the γ factor in the inequality breaks symmetry in the system, meaning that the orientation of the state with respect to the measurements does not allow either of Alice or Bob's measurements to be along the z-direction, and achieve the value v with the state $|\Phi(\beta)\rangle$. If none of the measurements are along the z-axis then the marginal randomness must be greater than $\cos^2 \beta$. Therefore, it is necessary to optimize the randomness over the state angle β and the angle of orientation of Alice's measurement to the z-axis, in order to improve on the bound given in [3] for the CHSH inequality. However this is not taken into account in the presented analysis.

---

[1] S. Vadhan, *Pseudorandomness*, vol. 7 of *Foundations and Trends in Theoretical Computer Science* (Now Publishers, 2012).
[2] R. Colbeck, Ph.D. thesis, University of Cambridge (2006), arXiv:0911.3814.
[3] 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., Nature **464**, 1021 (2010).
[4] S. Pironio and S. Massar, Phys. Rev. A **87**, 012336 (2013).
[5] S. Fehr, R. Gelles, and C. Schaffner, Phys. Rev. A **87**, 012335 (2013).
[6] U. Vazirani and T. Vidick (2012), arXiv:1111.6054.
[7] R. Colbeck and R. Renner, Nature Physics **8**, 450 (2012).
[8] R. Gallego, L. Masanes, G. D. L. Torre, C. Dhara, L. Aolita, and A. Acín, Nat. Commun. **4**, 2654 (2013).
[9] P. Mironowicz and M. Pawłowski (2013), arXiv:1301.7722.
[10] P. Mironowicz and M. Pawłowski, Phys. Rev. A **88**, 032319 (2013).

[11] B. G. Christensen, K. T. McCusker, J. Altepeter, B. Calkins, T. Gerrits, A. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, et al., Phys. Rev. Lett. **111**, 130406 (2013).

[12] M. J. W. Hall, Phys. Rev. A **84**, 022102 (2011).

[13] J. Barrett and N. Gisin, Phys. Rev. Lett. **106**, 100406 (2011).

[14] D. E. Koh, M. J. W. Hall, Setiawan, J. E. Pope, C. Marletto, A. Kay, V. Scarani, and A. Ekert, Phys. Rev. Lett. **109**, 160404 (2012).

[15] T. Le, L. Sheridan, and V. Scarani, Phys. Rev. A **87**, 062121 (2013).

[16] R. Gill, Mathematical Statistics and Applications: Festschrift for Constance van Eeden. Eds: M. Moore, S. Froda and C. Léger. IMS Lecture Notes – Monograph Series **42**, 133 (2003).

[17] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu, Phys. Rev. A **66**, 042111 (2002).

[18] A. Acín, S. Massar, and S. Pironio, Phys. Rev. Lett. **108**, 100402 (2012).

[19] C. Dhara, G. de la Torre, and A. Acín, Phys. Rev. Lett. **112**, 100402 (2013).

[20] M. Navascues, S. Pironio, and A. Acin, Phys. Rev. Lett. **98**, 010401 (2007).

[21] M. Navascues, S. Pironio, and A. Acin, New Journal of Physics **10**, 073013 (2008).

[22] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov, and C. Kurtsiefer, Phys. Rev. Lett. **107**, 170404 (2011).

[23] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, et al., Nature **497**, 227 (2013).

[24] P. H. Eberhardt, Phys. Rev. A **47**, R747 (1993).

[25] J. Sturm, Optimization Methods and Software **11–12**, 625 (1999), version 1.05, URL `http://fewcal.kub.nl/sturm`.

[26] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, Phys. Rev. Lett. **111**, 030501 (2013).

[27] URL `http://www.faacets.com/db/solved`.

[28] O. Nieto-Silleras, S. Pironio, and J. Silman, New. J. Phys. **16**, 013035 (2014).

[29] S. Boyd and S. Vandenberghe, *Convex optimization* (Cambridge University Press, 2004).

[30] R. Horodecki, P. Horodecki, and M. Horodecki, Phys. Lett. A **200**, 340 (1995).

[31] These curves were actually computed using an SDP relaxation as described previously, but they were checked to be optimal.

[32] Note that a similar claim is presented in [23], but without enough details to allow for a quantification of the randomness created during the experiment. We thus restrict our analysis to the data reported in [11].

[33] Note that in [11] authors consider extracting randomness from the outcomes observed by one party rather than from joint outcomes. For the small CHSH violation observed here, the amount of randomness present in the joint outcomes is however sensibly similar (see [3]).