Unbounded randomness certification from one pair of qubits using sequences of measurements

Florian J. Curchod,[1] Markus Johansson,[1] Remigiusz Augusiak,[1] Matty J. Hoban,[2] Peter Wittek,[1] and Antonio Acín[1,3]

[1]*ICFO–Institut de Ciencies Fotoniques, 08860 Castelldefels (Barcelona), Spain.*
[2]*Department of Computer Science, University of Oxford, Oxford OX1 3QD, United Kingdom.*
[3]*ICREA–Institució Catalana de Recerca i Estudis Avançats, E–08010 Barcelona, Spain*

(Dated: July 27, 2018)

Unpredictability, or randomness, of the outcomes of measurements made on an entangled state can be *certified* provided that the statistics violate a Bell inequality. In the standard Bell scenario, each party performs a single measurement on his share of the system before receiving a fresh one from the source. In this scenario, recent work proved an upper bound of $2log_2(d)$ on the amount of random bits certifiable from a pair of qudits. In our work, it is shown that this fundamental limitation can be overcome using sequences of (non projective) measurements on the same system. More precisely, we prove that one can certify *any* amount of random bits from a pair of qubits in a pure state as resource, even if arbitrarily weakly entangled. Moreover, this certification is achieved by near-maximal violation of a particular Bell inequality for each measurement in the sequence.

Bell's theorem [1] has shown that the predictions of quantum mechanics demonstrate non-locality. That is, they cannot be described by a theory in which there are objective properties of a system prior to measurement that satisfy special relativity (sometimes referred to as "local realism"). If one requires special relativity to be satisfied at the operational level then it can be shown that the outcomes of measurements demonstrating non-locality must be unpredictable [1–3]. This unpredictability, or randomness, is not the result of ignorance about the system preparation but is *intrinsic* to the theory.

Although this connection between quantum non-locality (via Bell's theorem) and the existence of intrinsic randomness is well know [1, 3, 4] it was only analysed in a quantitative way recently [5, 6]. It was shown how to use non locality (probability distributions that violate a Bell inequality) to *certify* the impredictability of the outcomes of certain physical processes. This was termed *device independent randomness certification*, as the certification only relies on the statistical properties of the outcomes and not on how they where produced. The development of device-independent randomness expansion [5, 7] and of randomness amplification [8, 9] then followed. For example, quantum theory is somehow an optimal theory in that it exhibits both non-locality and maximal randomness whereas other potential theories (more general than quantum theory) may have more non-locality but less randomness [10].

Entanglement is a necessary resource for quantum non-locality, that in turn is required for randomness certification. It is thus crucial to understand qualitatively and quantitatively how these three fundamental quantities relate one to another. In our work, we focus on asking how much certifiable randomness can be obtained from a single entangled state as a resource. Progress has been made in this direction for entangled states shared between two parties, Alice (A) and Bob (B), in the standard scenario where each party makes a single measurement on his share of the system and then discards it. An argument adapted from Ref. [11] shows that either of the two parties, Alice or Bob can certify at most $2\log_2(d)$ bits of randomness [12], where d is the dimension of the local Hilbert space the state lives in. This demonstrates a fundamental limitation for device-independent randomness certification in this standard scenario. The goal of our work is to show that this limitation on the amount of certifiable random bits from one quantum state can be lifted. To do this we will work with sequences of measurements on the same system.

In this sequential scenario, one (or both) of the parties makes a measurement but this measurement does not completely destroy the entanglement of the system and a new measurement can be made on the post-measurement state instead of discarding it. In this way, we will show that by increasing the number of consecutive measurements made on the state it is possible to certify the production of *any* amount of random bits.

To gain intuition, consider the following set-up where the functioning of a device can be entirely trusted. The device consists of a quantum state prepared in the Pauli-Z, or σ_z eigenstate $|0\rangle$ followed by a measurement in the Pauli-X, or σ_x basis $\{|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}, |-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}\}$. The outcome of this measurement is random and if the device then makes another measurement on the state, this time in the Pauli-Z basis this gives yet another random outcome. In this fashion of alternating between the two orthogonal bases, one can potentially obtain an unbounded number of random bits from one qubit. The limitation of this procedure for producing random numbers is that one cannot distinguish this device from a classical one with pre-programmed outcomes - a *local model* for the outcomes - if one doesn't trust the functioning of the device completely.

Clearly we cannot *certify* any randomness from a single system (in a device-independent manner) as in the above example, since one needs non locality and entanglement for this purpose. But is it possible to build a scheme, that exploits non-locality and makes use of this idea of measuring the state repeatedly, to overcome the bound on the amount of certifiable randomness that one

can obtain from a single entangled quantum system? To do so, one of the challenges is to come up with non-destructive measurements that still produce non-locality but retain some entanglement in the post-measurement state. In this way, the state can still be used as a resource for subsequent measurements. Bell tests with sequences of measurements have received less attention than the standard ones with single measurement in the literature despite the novel features in this scenario [17, 18]. In our work we show that they prove useful in the task of randomness certification, which also provides another example [add ref] where general measurements (POVMs) can overcome limitations of projective ones.

The main result in this work is to show that the bound of $2\log_2(d)$ random bits in the standard scenario can be overcome. In fact, we can potentially certify an *unbounded* amount of randomness, even using two qubits. More precisely, we describe a scheme where any number m of random bits to be certified using a sequence of $n > m$ consecutive measurements on the same state. Moreover, this unbounded randomness is certified by a near-maximal violation of a particular Bell inequality for each measurement in the sequence.

The sequential measurements scenario.– Before presenting our results, let us introduce the scenario we work in. We carry over many of the features from the standard scenario except now we allow party Bob to make multiple dichotomic measurements in a sequence on his share of the state. One can visualise this as in Fig.1 where Bob is split up into several Bobs, each one corresponding to a measurement made on the state and is labelled by B_i, $i \in \{1, 2, .., n\}$, where n is the total number of measurements made in the sequence. Each B_i makes one measurement and the post-measurement state is sent to B_{i+1} [24]. We organize the Bobs such that B_i is doing his measurement *before* B_j for $i < j$. Thus in principle B_j can receive the information about the inputs and outputs of previous measurements B_i for all $i < j$.

To put the above scenario in the setting of *non-local guessing games* (e.g. Refs. [10, 12–14]), let us consider an additional adversary Eve (E) that is in possession of a quantum system potentially correlated to the one of A and B. The global state is denoted ρ_{ABE}. We assume that at each round of the experiment E is the one preparing the state ρ_{ABE} and distributes $\rho_{AB} = Tr_E(\rho_{ABE})$ to A and B. This state will be used to make the measurements in the sequence and the aim of E is to try to guess B's outcomes by using measurements on her share of the state ρ_{ABE}. A and B, having no knowledge about the state or the real measurements made on it, see their respective devices as black boxes that receive some classical input $x \in \{0, 1\}$ and $y_1, y_2, .., y_n \equiv \vec{y}$, $y_i \in \{0, 1\}$, respectively and that generate a classical output $a \in \{\pm 1\}$ and $b_1, b_2, .., b_n \equiv \vec{b}$, $b_i \in \{\pm 1\}$, respectively (see Fig.1). They generate statistics from multiple runs of the experiment to obtain the observed probability distribution P_{obs} with elements $p_{\text{obs}}(a, \vec{b}|x, \vec{y})$. This distribution P_{obs} lives inside the set of quantum correlations $\mathcal{Q}$ obtained from

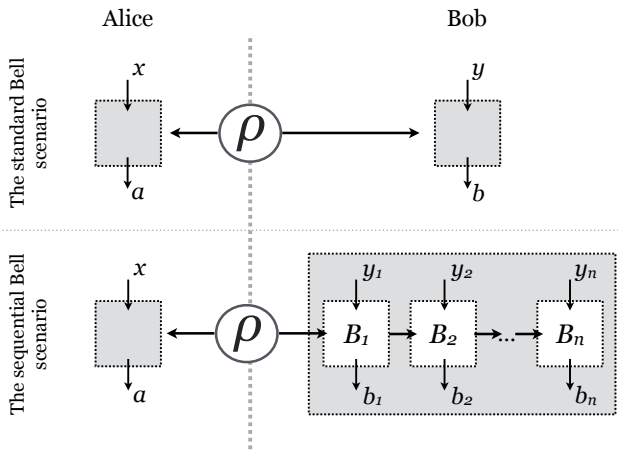

FIG. 1: The standard scenario, where parties A and B make a single quantum measurement on their share of the state and discard it versus the sequential scenario where the second party B makes multiple measurements on his share.

measurements on quantum states in a sequence as we described. This set is convex and thus can be described in terms of its extreme points which will be denoted P_{ext} such that $P_{\text{obs}} = \sum_{\text{ext}} q_{\text{ext}} P_{\text{ext}}$ and $\sum_{\text{ext}} q_{\text{ext}} = 1$ and every $q_{\text{ext}} \geq 0$.

From studying the outcomes statistics *only* we can bound E's predictive power by allowing her to have complete knowledge of how P_{obs} is decomposed into extreme points, i.e. she knows the probability distribution q_{ext} over extreme points P_{ext}. This predictive power is quantified via the *device-independent guessing probability* (DIGP) [add ref for this] where we fix the particular input string $y_1^0, y_2^0, .., y_n^0 \equiv \vec{y}^0$ for which E has to guess the outputs $\vec{b}$. The DIGP, denoted by $G(\vec{y}^0, P_{\text{obs}})$, is then calculated as the optimal solution to the following optimization problem [10, 14]:

$$G(\vec{y}^0, P_{\text{obs}}) = \max_{\{\vec{b}, q_{\text{ext}}, P_{\text{ext}}\}} \sum_{\text{ext}} q_{\text{ext}} p_{\text{ext}}(\vec{b}|\vec{y}^0)$$

subject to:

$$p_{\text{ext}}(\vec{b}|\vec{y}^0) = \sum_a p_{\text{ext}}(a, \vec{b}|x, \vec{y}^0), \qquad \forall x \tag{1}$$

$$P_{\text{obs}} = \sum_{\text{ext}} q_{\text{ext}} P_{\text{ext}}, \qquad P_{\text{ext}} \in \mathcal{Q}. \tag{2}$$

One can recover the formalism of the standard scenario by simply considering that $\vec{b} = b$ and $\vec{y}^{(0)} = y^{(0)}$. To quantify the amount of bits of randomness that is certified, we use the *min entropy* $H(\vec{y}^0, P_{\text{obs}}) = -\log_2 G(\vec{y}^0, P_{\text{obs}})$ which returns m bits of randomness if $G(\vec{y}^0, P_{\text{obs}}) = 2^{-m}$. The amount of bits of randomness quantified this way is the figure of merit in this workand our goal is to obtain as many bits as possible from a single system. [Shall we add some intuition about what this DIGP means ? Something like it corresponds to giving E all the power on the realisation (know the state, know

the meas, ie know the P_ext), but to reproduce the correlations. A and B base the randomness proof on the statistics, whatever the actual realisation was. Just to gain intuition, very mathematical here, or is it clear you think ?]

Finally, let us note that the problem in (2) can be further relaxed to an optimization where instead of insisting on $P_{\text{obs}} = \sum_{\text{ext}} q_{\text{ext}} P_{\text{ext}}$ (2), we only impose the constraint that the observed statistics P_{obs} give a particular Bell inequality violation [5]. Again add some kind of intuition about the new "constraint"? What does it mean for E The optimal solution to this new problem will be an upper bound to the optimal solution of (2). Crucially, this relaxation still gives good bounds as we will show in the following discussion.

The ingredients.– Alice and Bob share the pure two-qubit state

$$|\psi(\theta)\rangle = \cos(\theta)|00\rangle + \sin(\theta)|11\rangle \tag{3}$$

that for all $\theta \in]0, \frac{\pi}{2}[$ is entangled. Any pure (entangled) two qubit state can be written in this form up to some local change of basis.

In Ref. [13], a family of Bell inequalities was introduced:

$$I_\theta = \beta\langle \mathbb{B}_0\rangle + \langle \mathbb{A}_0\mathbb{B}_0\rangle + \langle \mathbb{A}_1\mathbb{B}_0\rangle + \langle \mathbb{A}_0\mathbb{B}_1\rangle - \langle \mathbb{A}_1\mathbb{B}_1\rangle \tag{4}$$

where $\beta = \frac{2\cos(2\theta)}{\sqrt{1+\sin^2(2\theta)}}$, $\langle \mathbb{B}_y\rangle = p(b=+1|y) - p(b=-1|y)$ and $\langle \mathbb{A}_x\mathbb{B}_y\rangle = p(a=b|xy) - p(a\neq b|xy)$ for x,$y \in \{0,1\}$. The maximal quantum value of the inequality $I_\theta = I_\theta^{max} = 2\sqrt{2}\sqrt{1+\frac{\beta^2}{4}}$ is obtained by measuring the state (3) with:

$$\begin{aligned} \mathbb{A}_0 &= \cos(\mu)\sigma_z + \sin(\mu)\sigma_x, & \mathbb{B}_0 &= \sigma_z, \\ \mathbb{A}_1 &= \cos(\mu)\sigma_z - \sin(\mu)\sigma_x, & \mathbb{B}_1 &= \sigma_x, \end{aligned} \tag{5}$$

where $\tan(\mu) = \sin(2\theta)$.

This family of inequalities has the following two useful properties : (i) it is maximally violated by the state $|\psi(\theta)\rangle$ with angle θ and (ii) when maximally violated, this inequality certifies one bit of local randomness on Bob's side for his second measurement choice $y^0 = 1$: $G(y^0 = 1, P_{\text{obs}}^{\max}) = \frac{1}{2}$ [13]. These observations are possible because the maximal violation is *uniquely* achieved by the probability distribution $P_{\text{obs}}^{\max}$ that arises from the previously-described state and measurements (3) and (5). Therefore, for the maximal violation, $P_{\text{obs}}^{\max} = P_{\text{ext}}$ in (2) and the guessing probability for input choice $y^0 = 1$ is $\frac{1}{2}$.

However, we may not in general get correlations that maximally violate our Bell inequality but give a violation that is only close to maximal. In the appendices [I cannot make the reference to the appendices, for some reason] we show how to make conclusions about the guessing probability for non maximal violations. In particular, we show that for *any* Bell inequality with a unique point of maximal violation, the guessing probability is a continuous function of the value of the inequality close to the maximal violation. This implies in the particular case we are studying that:

$$I_\theta \to I_\theta^{max} \quad \Rightarrow \quad G(y^0 = 1, P_{\text{obs}}) \to \frac{1}{2}. \tag{6}$$

In appendix [I cannot make the reference to the appendices, for some reason], we also provide a numerical upper bound on the guessing probability $G(y^0 = 1, P_{obs})$ by a concave function of the value of I_θ.

So far we have reviewed and rigorously expanded on the results in Ref. [13] showing how much randomness can be obtained from a pure entangled two-qubit state for dichotomic measurements. As mentioned, we can obtain more local randomness by considering measurements with more outcomes but it is ultimately limited to two bits [12, 13]. To go beyond this restrictive limit we work in a more general Bell scenario that allows for sequences of measurements (see Fig.1) made on the state.

If B_1 performs a projective measurement on $|\psi(\theta)\rangle$ (3) the post-measurement state now shared between Alice and B_2 is separable and thus useless for randomness production. Consequently, one needs to consider non-projective POVMs to retain some entanglement in the system for the subsequent measurements. For this purpose, let us introduce the following two-outcome quantum measurements (written in the formalism of Kraus operators):

$$M_{\pm 1} = \cos(\xi)|\pm\rangle\langle\pm| + \sin(\xi)|\mp\rangle\langle\mp| \tag{7}$$

corresponding to the two outcomes $\{\pm 1\}$. To gain intuition, these measurements can be expressed as the following observable:

$$\hat{\sigma}_x(\xi) = \cos(2\xi)\cdot\sigma_x = M_{+1}^\dagger M_{+1} - M_{-1}^\dagger M_{-1}$$

[I cannot add the equation number in this equation, no idea why...] This observable can be understood as a noisy version of the projective measurement of the observable σ_x. One can check that the measurement of $\hat{\sigma}_x(\xi)$ varies from being projective (for $\xi = 0$) to being non-interacting (for $\xi = \frac{\pi}{4}$). Also, one can verify that measuring an entangled state (3) for $\xi \in]0, \frac{\pi}{4}]$ (non-projective measurement) the post-measurement state still retains some entanglement, irrespectively of the outcome. Therefore, by tuning the parameter ξ we are able to vary the destruction of the entanglement of the state at the gain of extracting information from it (cf. Ref. [19]). Intuitively, the closer to being a projective measurement, the lower the entanglement in the post-measurement state, but the bigger the violation of the initial Bell inequality. On the other hand, one can leave a lot of entanglement in the state by barely interacting with it but at the cost of a lower Bell inequality violation.

A scheme for unbounded randomness certification.–We now combine the previous observations to demonstrate our main result. First, let us recall that, as shown in [13], one can obtain one bit of randomness from any pure entangled two qubit state, irrespectively of the amount of entanglement in it. Moreover, one can verify that approximately one random bit can be certified if the measurements are close to the ones in (5) (in the sense that $\hat{\sigma}_x(\xi) \to \sigma_x$ for $\mathbb{B}_1$ (5)) since I_θ is then close to I_θ^{max} (6). Second, the measurement in Eq. (7) is only close-to-projective for ξ close to zero and leaves entanglement in the post-measurement state between Alice and Bob which is thus still useful for randomness certification. By repeated use of these two properties we can certify the production an unbounded amount of random bits from a single pair of entangled qubits. We now formally describe this process in which Alice makes a single measurement on her share of the state, whereas Bob makes a sequence of n measurements on his.

Each B_i chooses between measurements of σ_z and $\hat{\sigma}_x(\xi_i)$ for inputs $y_i = 0$ and $y_i = 1$ respectively, with outcomes $b_i \in \{\pm 1\}$. The parameter ξ_i is fixed before the beginning of the experiment. The initial entangled state shared between Alice and Bob, before B_1's measurement, is $|\psi^{(1)}(\theta_1)\rangle$ ((3) with $\theta = \theta_1$). If the first non-projective measurement of the operator $\hat{\sigma}_x(\xi_1)$ is made by B_1 on the initial state $|\psi^{(1)}(\theta_1)\rangle$, the post-measurement state is of the form

$$|\psi^{(2)}_{b_1}(\theta_1, \xi_1)\rangle = U_A^{b_1}(\theta_1, \xi_1) \otimes V_B^{b_1}(\theta_1, \xi_1)(c|00\rangle + s|11\rangle), \tag{8}$$

where $c = \cos(\theta_{b_1}(\theta_1, \xi_1))$ and $s = \sin(\theta_{b_1}(\theta_1, \xi_1))$ and the two unitaries, $U_A^{b_1}(\theta_1, \xi_1)$ and $V_B^{b_1}(\theta_1, \xi_1)$, and angle $\theta_{b_1}(\theta_1, \xi_i) \in]0; \frac{\pi}{4}]$ depend on the first outcome b_1 and the angles θ_1 and ξ_1.

After his measurement, B_1 applies the unitary $(V_B^{b_1})^\dagger$, conditioned on his outcome b_1, on the post-measurement state going to B_2. This allows B_2 to use the same two measurements $\hat{\sigma}(\xi_2)$ and σ_z independently of the outcome b_1 since the unitary $(V_B^{b_1})$ is cancelled in (8). This last procedure will be applied by each B_i after his measurement, before sending the post-measurement state to the next B_{i+1}. If the system passed through *only* the non-projective measurements, the state received by B_i can be one of 2^{i-1} potential states, depending on all of the previous B_j's ($j < i$) outcomes (one for each combination $\vec{b}_{i-1} \equiv (b_1, b_2, .., b_{i-1})$ of outcomes obtained by the previous B_j, these can be computed *before* the beginning of the experiment). One of these states can be written as:

$$|\psi^{(i)}_{\vec{b}_{i-1}}\rangle = U_A^{\vec{b}_{i-1}} \otimes \mathbb{1}_B \left(\cos(\theta_{\vec{b}_{i-1}})|00\rangle + \sin(\theta_{\vec{b}_{i-1}})|11\rangle\right), \tag{9}$$

where the angles $\theta_{\vec{b}_{i-1}}$ and the matrix $U_A^{\vec{b}_{i-1}}$ both depend on the outcomes $\vec{b}_{i-1}$, on the initial angle θ_1 and the angles ξ_j of the previous B_j's with $j < i$. In the notation, we will always omit the dependence on the angles θ_1 and $\xi_1, \xi_2, .., \xi_j$ since these are fixed *before* the beginning of the experiment. For each of these different potential states with angle $\theta_{\vec{b}_{i-1}}$, Alice adds two measurements to her input choices, where for $k \in \{0, 1\}$, these are measurements of the observables $\mathbb{A}_k^{\vec{b}_{i-1}}$ which are defined as

$$U_A^{\vec{b}_{i-1}} \left(\cos(\mu_{\vec{b}_{i-1}})\sigma_z + (-1)^k \sin(\mu_{\vec{b}_{i-1}})\sigma_x\right) (U_A^{\vec{b}_{i-1}})^\dagger, \tag{10}$$

where $\tan(\mu_{\vec{b}_{i-1}}) = \sin(2\theta_{\vec{b}_{i-1}})$, depending on the specific state $|\psi^{(i)}_{\vec{b}_{i-1}}\rangle$ (9).

We are now ready to describe how the scheme certifies randomness. The measurement operator $\hat{\sigma}_x(\xi_i)$ can be made arbitrarily close to σ_x by choosing ξ_i sufficiently small. This brings the outcome statistics for measurements $\hat{\sigma}_x(\xi_i), \sigma_z$ on Bob's side and $\mathbb{A}_0^{\vec{b}_{i-1}}, \mathbb{A}_1^{\vec{b}_{i-1}}$ on Alice's side on the state in Eq. (9), arbitrarily close to the statistics for the measurements in Eq. (5) and a state of the form in Eq. (3), for $\theta = \theta_{\vec{b}_{i-1}}$. Therefore, the inequality $I_{\theta_{\vec{b}_{i-1}}}$ for Alice and B_i as defined in (4) can be made arbitrarily close to its maximal violation. This in turn guarantees that the guessing probability, $G(y_i^0 = 1, P_{obs})$ can be made arbitrarily close to $1/2$. Note that this guessing probability does not only describe the instances when Alice chooses the measurements $\mathbb{A}_j^{\vec{b}_{i-1}}$. Since Eve does not know Alice's measurement choices in advance she cannot use a strategy that gives higher predictive power for the instances when Alice chooses other measurements. Finally, by making $G(y_i^0 = 1, P_{obs})$ sufficiently close to $1/2$ for each i (by choosing each ξ_i sufficiently close to 0) the DIGP $G(y_1^0, y_2^0, .., y_n^0, P_{obs})$ can be made arbitrarily close to 2^{-n} (see appendix for a proof[again, I cannot make the reference to the appendices, for some reason]).

At the end, Bob can produce m random bits by a suitably chosen sequence $\hat{\sigma}_x(\xi_i)$, $i \in \{1, 2, .., n\}$, of $n > m$ measurements. The certification only requires that each B_i occasionally chooses the projective measurement σ_z so that the whole statistics can be obtained. Note that Bob can choose σ_z with probability γ_i and $\hat{\sigma}_x(\xi_i)$ with probability $1 - \gamma_i$ for γ_i as close to zero as he wants.

To summarise the idea of the scheme, the post-measurement state after B_{i-1}, given a sequence of non-projective measurements, is of the form of (9). With some probability Alice chooses measurements $\mathbb{A}_k^{\vec{b}_{i-1}}$ which give outcome statistics that allow randomness certification of the ith bit using the $I_{\theta_{\vec{b}_{i-1}}}$ inequality. Therefore, we can certify randomness for each measurement B_i in the sequence at the expense of increasing the number of measurements that Alice chooses from. Finally, also remark that the value of *each* inequality $I_{\theta_{\vec{b}_{i-1}}}$ between each B_i and A can be made as close as wanted to the maximal value $I^{\max}_{\theta_{\vec{b}_{i-1}}}$.

Conclusion.– We have presented a scheme for certifying an unbounded amount of random bits from a single pair of entangled qubits where one of the qubits is subjected to a sequence of measurements. The mea-

surements do not completely destroy the entanglement but map the state to another pure entangled two-qubit state (with reduced entanglement). We obtain this certified randomness at the expense of exponentially increasing the number of measurements that Alice needs to make ($\sum_{i=1}^{n} 2^i$ measurement choices for n measurements in the sequence), thus providing an immediate obstruction to constructing a device-independent randomness *expansion* protocol [5, 7]. Another obstruction is the sensitivity to noise in the state and measurements that may inhibit a fault-tolerant protocol.

Our main result made use of the fact that every measurement in Bob's sequence generated an almost-maximally non-local output distribution (in the sense of violating some Bell inequality almost maximally). A similar study was made in Ref. [19], where a sequence of non-local correlations was already obtained from a single pair of qubits. Nevertheless, it was still open to know whether it was possible to get a sequence of (arbitrarily close to) maximally non-local correlations, and our results show that this is possible, a property that could be of further use for many other device-independent quantum information tasks.

Finally, on a more fundamental level, our results offer new insights in the relation between entanglement, non-locality and randomness. We showed that a single pair of pure entangled qubits is a potentially unbounded source of certifiable random bits (and of non-local correlations) if we perform sequences of measurements on it.

Acknowledgements.– F.J.C and M.J. acknowledge support from the John Templeton Foundation, the Generalitat de Catalunya (SGR 875) and the Spanish Project FOQUS. M.J. also acknowledges support from the Marie Curie COFUND action through the ICFOnest program. [Add or send me your stuff to acknowledge please :-)])

[1] J. S. Bell, Physics (NY) **1** (3), 195-200 (1964).
[2] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[3] Ll. Masanes, A. Acín, and N. Gisin, Phys. Rev. A. **73**, 012112 (2006).
[4] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[5] S. Pironio *et al.*, Nature (London) **464**, 1021 (2010).
[6] R. Colbeck, PhD Thesis University of Cambridge, arXiv:0911.3814 (2006).
[7] U. Vazirani, and T. Vidick, Phil. Trans. R. Soc. A **370**, 34323448 (2012).
[8] R. Colbeck, and R. Renner, Nature Physics **8**, 450-453 (2012).
[9] R. Gallego, L. Masanes, G. De La Torre, C. Dhara, L. Aolita, and A. Acín, Nature Communications **4**, 2654 (2013).
[10] G. de la Torre, M. J. Hoban, C. Dhara, G. Prettico, A. Acín, Phys. Rev. Lett. **114**, 160502 (2015).
[11] G. M. DAriano, P. Lo Presti and P. Perinotti, J. Phys. A: Math. Gen. **38** 5979, (2005).
[12] A. Acín, S. Pironio, T. Vértesi and P. Wittek, arXiv:1505.03837 (2015).
[13] A. Acín, S. Massar and S. Pironio, Phys. Rev. Lett. **108**, 100402 (2012).
[14] O. Nieto-Silleras, S. Pironio and J. Silman, New J. Phys. **16**, 013035 (2014).
[15] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[16] T. H. Yang and M. Navascués, Phys. Rev. A **87**, 050102(R) (2013).
[17] R. Gallego, L. E. Würflinger, R. Chaves, A. Acín, and M. Navascués, New J. Phys. **16** 033037 (2014).
[18] S. Popescu, Phys. Rev. Lett. **74**, 2619 (1995).
[19] R. Silva, N. Gisin, Y. Guryanova and S. Popescu, arXiv:1408.2272 [quant-ph] (2014).
[20] M. Coudron and H. Yuen, Proc. ACM Symp. on Theory of Comp. (STOC) (2014).
[21] K. -M. Chung, Y. Shi and X. Wu, arXiv:1402.4797 [quant-ph] (2014).
[22] R. T. Rockafellar, *Convex analysis* Princeton Press, (1970),Theorem 10.1
[23] O. Bucicovschi and J. Lebl, J. Convex Anal. **20**, 1113 (2013).
[24] For the sake of brevity we discuss systems being sent between different Bobs but we allow the possibility that there is one physical box and a sequence of inputs are made to this one box where, for each input, an output is obtained.

Appendices

The guessing probability

We start our appendices with the following discussion, which is a summary of the work done in deriving the device-independent guessing probability (DIGP) [5, 10, 13, 14]. A probability distribution that is the outcome distribution for some measurement on a quantum state is called a quantum distribution. For example, a distribution P with elements $p(ab|xy)$ is quantum if there exist at least one positive semi-definite hermitian unit trace matrix ρ and at least one set of positive semi-definite hermitian matrices X_i, Y_i satisfying $\sum_i X_i = 1$ and $\sum_i Y_i = 1$ such that $p(ab|xy) = Tr(X_a \otimes Y_b \rho)$. We will often abuse notation and refer to a distribution by its elements $p(ab|xy)$ when there is no confusion in doing so.

The set $\mathcal{Q}$ of quantum distributions is closed and convex and a distribution in $\mathcal{Q}$ that cannot be decomposed as a convex combination of other distributions is called *extremal* in $\mathcal{Q}$. For a non-extremal distribution $P(ab|xy)$ there is in general more than one possible convex decomposition.

A non-extremal distribution $p(ab|xy)$ with a convex decomposition $p(ab|xy) = \sum_\lambda q_\lambda p_\lambda(ab|xy)$ can be constructed by sampling the different distributions $p_\lambda(ab|xy)$ with probability q_λ. In this case knowledge about the convex decomposition chosen changes the ability of an eavesdropper to correctly guess the outcomes a and/or b.

Without knowledge of the decomposition, or for extremal distributions, the probability of correctly guessing the outcome of measurement y^0 is $\max_b p(b|y^0)$, the probability of the most likely outcome. With knowledge of the decomposition $p(ab|xy) = \sum_\lambda q_\lambda p_\lambda(ab|xy)$, the probability is larger or equal to $\max_b p(b|y^0)$

$$\sum_\lambda q_\lambda \max_b p_\lambda(b|y^0) \geq \max_b \sum_\lambda q_\lambda p_\lambda(b|y^0) = \max_b p(b|y^0). \tag{11}$$

For a given observed non-extremal distribution it is possible that it was produced by an agent Eve that has larger predictive power than an agent which only observes the outcomes. The maximal probability for the agent Eve to correctly guess an outcome b of measurement y^0 given a distribution $p(ab|xy)$ and a free choice of decomposition is the DIGP $G(y^0, P_{\text{obs}})$

$$G(y^0, P_{\text{obs}}) = \max_{q_\lambda, p_\lambda(ab|xy)} \sum_\lambda q_\lambda \max_b p_\lambda(b|y^0). \tag{12}$$

where λ is labelling the convex decompositions of $p_{\text{obs}}(ab|xy)$ in terms of extremal distributions $p_\lambda(ab|xy)$. For any open interval of $\mathcal{Q}$ the function $G(y^0, P_{\text{obs}})$ is a concave function [5]. Therefore this kind of maximization is called a *concave roof* construction.

Continuity of the guessing probability in interior and extremal points of $\mathcal{Q}$

We want to show that the following propositions are true:

Proposition 1. *The function $G(y^0, P_{obs})$ on the set of quantum distributions $\mathcal{Q}$ is continuous in the interior of $\mathcal{Q}$.*

Proposition 2. *The function $G(y^0, P_{obs})$ is continuous in any extremal point of $\mathcal{Q}$.*

Proposition 1 is trivial. The guessing probability $G(y^0, P_{\text{obs}})$ is concave by definition and any concave function is continuous on an open subset of its domain [22]. In particular this means that $G(y^0, P_{\text{obs}})$ is continuous in the interior of $\mathcal{Q}$.

To address proposition 2 we consider the restriction $G(y^0, P_{\text{obs}})^{\partial\mathcal{Q}}$ of $G(y^0, P_{\text{obs}})$ to the boundary $\partial\mathcal{Q}$ of the quantum set. First we note that the function $G(y^0, P_{\text{obs}})^{\partial\mathcal{Q}}$ by definition is continuous on any open set of extremal points since $\max_b p(b|y)$ is a continuous function. Next we observe that the boundary $\partial\mathcal{Q}$ can be decomposed into a collection of open sets of extremal points and a collection $\{S_i\}$ of closed connected possibly overlapping sets where each set is the closure of a maximal open connected subset. A maximal open connected subset M of the non-extremal points is an open set such that any other open connected set of non-extremal points which contains M is M itself. Therefore, each set S_i is the convex hull of the set of extremal points in its closure.

Any closed set S_i has a boundary ∂S_i with the rest of $\partial\mathcal{Q}$ which can be decomposed in the same way into open sets of extremal points and closed connected sets S_{ij} that are closures of maximal open connected sets of non-extremal points. The boundary ∂S_{ij} of S_{ij} with the rest of ∂S_i is in turn decomposable in the same way.

Continuing this successive decomposition of the boundary $\partial\mathcal{Q}$ we will eventually reach sets $S_{ijk\ldots}$ that are one dimensional simplexes, or alternatively sets with only extremal points in the boundary. On sets of these two types

$G(y^0, P_{\text{obs}})$ is a continuous function. To see this we introduce the following terminology, and use a theorem from Ref. [23].

A function for which all discontinuities are such that the function takes the higher value at a closed set and the lower value at an open set is called *upper semi-continuous*.

The function $G(y^0, P_{\text{obs}})^S$ defined on a closed convex set S can be viewed as an extension of $G(y^0, P_{\text{obs}})^{\partial S}$ to the interior of S. This extension is called the *concave roof extension*.

Theorem 1. *Let C be a compact set and $K = co(C)$ be the convex hull of C. If $F : C \to \mathbb{R}$ is bounded, upper semi-continuous, and concave on C, then the concave roof extension $\hat{F} : K \to \mathbb{R}$ of F to K is upper semi-continuous [23].*

The guessing probability is bounded and concave by definition. If the boundary of S has only extremal points it follows that $G(y^0, P_{\text{obs}})^{\partial S}$ is continuous in ∂S and by theorem 1 $G(y^0, P_{\text{obs}})^S$ is upper semi-continuous on S. Moreover, since $G(y^0, P_{\text{obs}})^S$ is concave it cannot have an upper semi-continuous discontinuity between the boundary and the interior. If S is a one-dimensional simplex we can, if necessary, restrict the domain of the guessing probability to a one dimensional subspace and make the same argument.

Next we consider discontinuities between S and an open set of extremal points.

Lemma 1. *Any discontinuity of $G(y^0, P_{obs})$ between a closed set and an open set of extremal points is upper semi-continuous.*

Proof. If the boundary point of the closed set is extremal the $G(y^0, P_{\text{obs}})$ is continuous since $\max_b p(b|y^0)$ is continuous. Next consider a non-extremal boundary point of the closed set. $G(y^0, P_{\text{obs}})$ in the non-extremal point is always greater or equal to $\max_b P(b|y^0)$ by Eq. 11. Thus any discontinuity is upper semi-continuous. □

If there is a discontinuity of $G(y^0, P_{\text{obs}})$ on the boundary of S it is, by lemma 1 , upper semi-continuous and at a set of non-extremal points.

By repeated application of Theorem 1 and lemma 1 we can conclude that $G(y^0, P_{\text{obs}})^{\partial \mathcal{Q}}$ is upper semi-continuous on $\partial \mathcal{Q}$ and that $G(y^0, P_{\text{obs}})$ is upper semi-continuous on $\mathcal{Q}$. Since $G(y^0, P_{\text{obs}})$ is concave there cannot be an upper semi-continuous discontinuity between the boundary $\partial \mathcal{Q}$ and the interior of $\mathcal{Q}$. Thus the only discontinuities are between non-extremal points in closed subsets of $\partial \mathcal{Q}$ and extremal points in open subsets of $\partial \mathcal{Q}$.

Bounds on the guessing probability as a function of a Bell inequality: Continuity at a unique point of maximal violation

The guessing probability as a function on the space of probability distributions is not everywhere continuous. An example of this is that the family of Bell-inequalities described in of Ref. [13] certifies one bit of randomness for measurements on a state with arbitrarily little entanglement. The probability distribution corresponding to such a state and the measurements in Eq. 5 has $G(y^0, P_o bs) = 1/2$ and is at the same time arbitrarily close to a distribution corresponding to measurements on a product state with $G(y^0, P_o bs) = 1$, i.e., a distribution which can be prepared by a local deterministic procedure. There is thus a discontinuity where the guessing probability jumps from $1/2$ to 1. The key to understanding this discontinuity is that the local deterministic distribution is not extremal while the quantum distribution in the neighbouring point is extremal. As seen in Eq. 11, the guessing probability is given by different functions depending on whether a distribution can be decomposed into other distributions or not, i.e., if it is extremal or not. This means discontinuities can appear at the boundary between extremal points and non-extremal points.

We will now show that discontinuities can *only* appear at such boundaries between extremal and non-extremal points in the boundary $\partial \mathcal{Q}$ of the quantum set $\mathcal{Q}$. To do this we use the property of the guessing probability described in Eq. 11, together with some general properties of concave functions and in particular concave roof constructions.

We have described the guessing probability as a function on set of quantum distributions, but it is sometimes useful to consider it as a function of the violation of some given Bell inequality I. A Bell expression is a linear function on the space of distributions and the set of distributions for which it takes a given value t is a hyper-plane H_t. The different values of the Bell expression thus defines a family of parallel hyperplanes.

On each hyperplane H_t we can consider the restriction $G(y^0, P_{\text{obs}})_t$ of $G(y^0, P_{\text{obs}})$ to the intersection of H_t with $\mathcal{Q}$ and take its maximum $\max G(y^0, P_{\text{obs}})_t$ on this intersection. This maximum is the highest probability for Eve to guess the outcome of y^0 for any distribution $P \in \mathcal{Q}$ such that $I(P) = t$. The function $\max G(y^0, P_{\text{obs}})_t$ can have a discontinuity at $t = t_c$ only if H_{t_c} intersects with a point in $\mathcal{Q}$ at which $G(y^0, P_{\text{obs}})$ is discontinuous.

Let us consider a Bell expression I and its maximal value t_{max} on $\mathcal{Q}$. If the intersection of $H_{t_{max}}$ and $\mathcal{Q}$ is a single extremal point it follows from Propositions 1 and 2 that there is a $t_c \neq t_{max}$ such that for the range $t_c \leq t \leq t_{max}$ for which $\max G(y^0, P_{\text{obs}})_t$ is a continuous function of t.

If the intersection of $H_{t_{max}}$ and $\mathcal{Q}$ contains more than one extremal point it also contains a set of non-extremal points of $\partial\mathcal{Q}$ and $G(y^0, P_{\rm obs})$ could have a discontinuity between this set and an open set of extremal points. This discontinuity could lead to a discontinuity of the function $\max G(y^0, P_{\rm obs})_t$ at t_{max}.

Guessing probability for a sequence

Let us consider a sequence of measurements $\hat{\sigma}(\xi_i)$ chosen by Bob and denote $(\xi_1, \xi_2, \ldots, \xi_n) \equiv \vec{\xi}$. The convex decomposition of the observed outcome distribution that gives Eve optimal probability to correctly guess the sequence of outcomes $\vec{b}_n$ of the measurements $(y_1^0, y_2^0, \ldots, y_n^0) \equiv \vec{y}_n^0$ is a function of $\vec{\xi}$. The guessing probability $G(\vec{y}_n^0, P_{\rm obs})$ is thus given by

$$G(\vec{y}_n^0, P_{\rm obs}) = \sum_{\lambda_{\vec{\theta}}} q_{\lambda_{\vec{\xi}}} \max_{\vec{b}_n} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \ldots p_{\lambda_{\vec{\xi}}}(b_n|\vec{y}_n^0 \vec{b}_{n-1}). \tag{13}$$

where the extremal distributions $p_{\lambda_{\vec{\xi}}}(b_n|y_n \ldots)$ and weights $q_{\lambda_{\vec{\xi}}}$ of the optimal convex decomposition are functions of $\vec{\xi}$ as indicated by the index $\lambda_{\vec{\xi}}$. Let us assume that a term which appears in the convex combination is

$$q_{\lambda_{\vec{\xi}}} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \ldots p_{\lambda_{\vec{\xi}}}(b_n|\vec{y}_n^0 \vec{b}_{n-1}). \tag{14}$$

Thus we assume that it corresponds to the most probable sequence of outcomes $\vec{b}_n$ for a specific distribution indexed by $\lambda_{\vec{\xi}}$.

Given that Eve has chosen the optimal convex decomposition for guessing the outcomes of $\vec{y}_n^0$ we consider her probability of correctly guessing the outcome of y_m^0 for $1 \leq m \leq n$ given a particular sequence of previous outcomes $\vec{b}_{m-1}$. It is given by

$$\sum_{\lambda_{\vec{\xi}}} k_{\lambda_{\vec{\xi}}} \max_{b_m} p_{\lambda_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1}), \tag{15}$$

where $k_{\lambda_{\vec{\xi}}}$ is the probability that the distribution indexed by $\lambda_{\vec{\xi}}$ will be sampled given the sequence of previous outcomes $\vec{b}_{m-1}$

$$k_{\lambda_{\vec{\xi}}} = \frac{q_{\lambda_{\vec{\xi}}} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \ldots p_{\lambda_{\vec{\xi}}}(b_{m-1}|\vec{y}_{m-1}^0 \vec{b}_{m-2})}{\sum_{\lambda_{\vec{\xi}}} q_{\lambda_{\vec{\xi}}} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \ldots p_{\lambda_{\vec{\xi}}}(b_{m-1}|\vec{y}_{m-1}^0 \vec{b}_{m-2})}. \tag{16}$$

The probability in Eq. 15 is larger or equal to $1/2$ but is lower or equal to $G(y_m^0, P_{\rm obs})$, the maximal probability that Eve could guess the outcome of y_m^0 correctly given that she had chosen an optimal strategy for this and not the optimal strategy for guessing the outcomes of the sequence $\vec{y}_n^0$. Thus if $G(y_m^0, P_{\rm obs})$ is close to $1/2$ so is the expression in Eq. 15.

Arbitrarliy close to n random bits for n measurements

We want to prove that $G(\vec{y}_n^0, P_{\rm obs})$ can be made arbitrarily close to 2^{-n} by making $G(y_m^0, P_{\rm obs})$ sufficiently close to $1/2$ for each $1 \leq m \leq n$.

The proof relies on the fact that if a convex combination of a collection of numbers x_i equals a, i.e., $\sum_i k_i x_i = a$ where $\sum k_i = 1$, and if $x_i \geq a$ for each i, it follows that for every i either $k_i = 0$ or $x_i = a$.

From this follows that when $G(y_m^0, P_{\rm obs})$ is very close to $1/2$ either $\max_{b_m} p_{\lambda_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1})$ in Eq. 15 is very close to $1/2$ or $k_{\lambda_{\vec{\xi}}}$ is very close to zero for each $\lambda_{\vec{\xi}}$. To see this more clearly we construct the following bound

$$\begin{aligned} k_{\lambda_{\vec{\xi}}} \max_{b_m} p_{\lambda_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1}) &\leq G(y_m^0, P_{\rm obs}) - \sum_{\lambda' \neq \lambda} k_{\lambda'_{\vec{\xi}}} \max_{b_m} p_{\lambda'_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1}) \\ &\leq G(y_m^0, P_{\rm obs}) - 1/2(1 - k_{\lambda_{\vec{\xi}}}) \end{aligned}$$

where we used $\max_{b_m} p_{\lambda'_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1}) \geq 1/2$ for each $\lambda'_{\vec{\xi}}$ and $\sum_{\lambda' \neq \lambda} k_{\lambda'_{\vec{\xi}}} = 1 - k_{\lambda_{\vec{\xi}}}$. It follows that

$$G(y_m^0, P_{\rm obs}) - 1/2 \geq k_{\lambda_{\vec{\xi}}} [\max_{b_m} p_{\lambda_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1}) - 1/2],$$

and given Eq. (16) this implies

$$G(y_m^0, P_{\text{obs}}) - 1/2 \geq q_{\lambda_{\vec{\xi}}} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \dots p_{\lambda_{\vec{\xi}}}(b_{m-1}|\vec{y}_{m-1}^0 \vec{b}_{m-2})[\max_{b_m} p_{\lambda_{\vec{\xi}}}(b_m|\vec{y}_n^0 \vec{b}_{m-1}) - 1/2].$$

Thus for sufficiently small $G(y_m^0, P_{\text{obs}}) - 1/2$ either $\max_{b_m} p_{\lambda_{\vec{\xi}}}(b_m|\vec{y}_m^0 \vec{b}_{m-1}) - 1/2$ can be made arbitrarily small, or the probability $q_{\lambda_{\vec{\xi}}} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \dots p_{\lambda_{\vec{\xi}}}(b_{m-1}|\vec{y}_{m-1}^0 \vec{b}_{m-2})$ that the distribution labelled by $\lambda_{\vec{\xi}}$ is sampled when y_m^0 is measured is made arbitrarily small.

The argument can be made for any B_m. For B_1, this it follows that either $p_{\lambda_{\vec{\xi}}}(b_1|y_1^0)$ is made arbitrarily close to $1/2$ or $q_{\lambda_{\vec{\xi}}}$ is made arbitrarily close to 0. For B_2, it follows that either $p_{\lambda_{\vec{\xi}}}(b_2|y_2^0 y_1^0 b_1)$ is made arbitrarily close to $1/2$ or $q_{\lambda_{\vec{\xi}}} p_{\lambda_{\vec{\xi}}}(b_1|y_1^0)$ is made arbitrarily close to zero. Given the second option and that $p_{\lambda_{\vec{\xi}}}(b_1|y_1^0)$ is made arbitrarily close to $1/2$ it is implied that that $q_{\lambda(\vec{\xi})}$ is made arbitrarily close to 0. If on the other hand $p_{\lambda_{\vec{\xi}}}(b_1|y_1^0)$ is not very close to $1/2$ it follows that $q_{\lambda_{\vec{\xi}}}$ is made arbitrarily close to zero by the preceding argument.

By induction it is clear that either the term in Eq. 14 satisfies that $p_{\lambda_{\vec{\xi}}}(b_1|y_1^0) \dots p_{\lambda_{\vec{\xi}}}(b_n|\vec{y}_n^0 \vec{b}_{n-1})$ can be made arbitrarily close to 2^{-n} or alternatively $q_{\lambda_{\vec{\xi}}}$ is made arbitrarily small. Since the same is true for every $\lambda_{\vec{\xi}}$ in Eq. 13 it follows that $G(\vec{y}_n^0, P_{\text{obs}})$ can be made arbitrarily close to 2^{-n}.

Numerical bounds on the guessing probability

We start once again with the I_θ inequality (4):

$$I_\theta = \beta\langle \mathbb{B}_0 \rangle + \langle \mathbb{A}_0 \mathbb{B}_0 \rangle + \langle \mathbb{A}_1 \mathbb{B}_0 \rangle + \langle \mathbb{A}_0 \mathbb{B}_1 \rangle - \langle \mathbb{A}_1 \mathbb{B}_1 \rangle \tag{17}$$

with $\beta = \frac{2\cos(2\theta)}{\sqrt{1+\sin^2(2\theta)}}$. In [13] it was proved that the maximal quantum value $I_\theta^{\max}$ for this inequality was given by:

$$I_\theta^{\max} = \sqrt{2(4+\beta^2)} = 2\sqrt{2}\left(1 + \frac{\cos^2(2\theta)}{1+\sin^2(2\theta)}\right)^{\frac{1}{2}} \tag{18}$$

As the authors explained, in order to maximize the bound of the inequality, one can restrict the space of quantum states to pure two qubit states [13], meaning that the maximal value achievable with quantum measurements on quantum states can always be achieved with (two-outcomes projective) measurements on a pure, two-qubit (entangled) state. Fixing the basis, we can work with the state (3):

$$|\psi(\theta)\rangle = \cos(\theta)|00\rangle + \sin(\theta)|11\rangle$$

Given this form of the state and by parametrising the measurements of Alice and Bob, one can verify numerically that:

$$I_\theta^2 + (2-\beta)^2\langle \mathbb{B}_1 \rangle \leq 2(4+\beta^2) \tag{19}$$

in the range $\beta \in [0, 2[$, i.e. $\theta \in [0, \frac{\pi}{2}[$, the whole range of interest where the state is entangled. From (19), it is easy to obtain an upper bound on the expectation value:

$$|\langle \mathbb{B}_1 \rangle| \leq \frac{\sqrt{2(4+\beta^2) - I_\theta^2}}{(2-\beta)} = \frac{\sqrt{(I_\theta^{\max})^2 - I_\theta^2}}{(2-\beta)} \tag{20}$$

where one inserts the *observed* value of the inequality I_θ in this equation. In term of the guessing probability, this gives:

$$P_G^{I_\theta}(y=1) \leq \frac{1}{2} + \frac{\sqrt{(I_\theta^{\max})^2 - I_\theta^2}}{2(2-\beta)} = f(I_\theta) \tag{21}$$

For example, one can insert the maximal quantum value $I_\theta^{\max}$ (18) in (20) or in (21) and get that $\langle \mathbb{B}_1 \rangle = 0$ or $P_G^\theta(y=1) = \frac{1}{2}$, which coincides with the certification of one perfect local random bit for input $y_0 = 1$ on Bob's side for the maximal violation of I_θ. Our bound is thus tight at the maximal violation of the inequality. Since the

probability distribution of maximal violation is unique, the point is necessarily an extreme point [13], so we can directly use the observed guessing probability of the eavesdropper to bound its predictive power (as an extreme point allows only for one decomposition).

If we now want to use our function to bound the guessing probability *inside* the set (not only at the point of maximal violation), and following the arguments of [13], one can check that the function $f(I_\theta)$ bounding the P_G^θ (21) is a concave function of its variable I_θ:

$$\partial^2_{I_\theta}(f(I_\theta)) = -\frac{2(4+\beta^2)}{(2(4+\beta^2)-I_\theta^2)^{\frac{3}{2}}} < 0 \tag{22}$$

where we used that both the numerator and denominator are positive from $I_\theta \leq I_\theta^{\max} = \sqrt{2(4+\beta^2)}$ (18). The bound can thus be extended to the points that do not necessarily violate maximally the inequality, and our bound $f(I_\theta)$ can be used in our protocol for unbounded randomness certification from a single pair of qubits.

We will now give some graphs of this upper bound on $P_G^{I_\theta}(y=1)$ from the region where the state (3) is maximally entangled ($\theta = \frac{\pi}{4}$), reproducing the CHSH scenario, to the one where it is only slightly entangled ($\theta \to 0$). This should provide one with an intuition of how close quantitatively to the maximal violation of I_θ^{max} the observed bound I_θ should be in order to get close to one perfect local bit of randomness ($P_G^{I_\theta}(y=1) \to \frac{1}{2}$) for a state with a given angle θ.

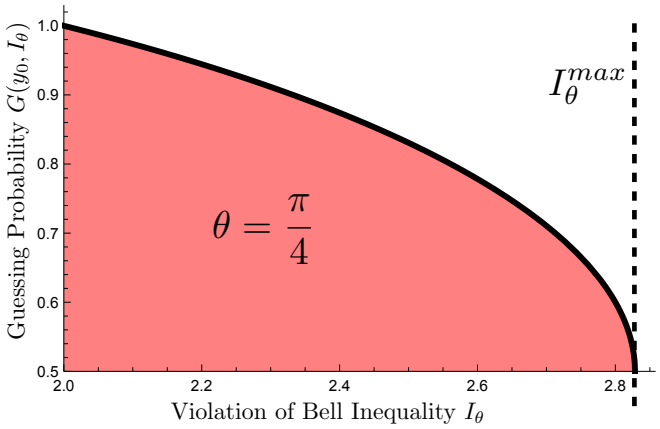

FIG. 2: The upper bound on the guessing probability in function of the violation of $I_{\theta=\frac{\pi}{4}}$ = CHSH, maximally violated by the maximally two qubit entangled state $\theta = \frac{\pi}{4}$ in (3).

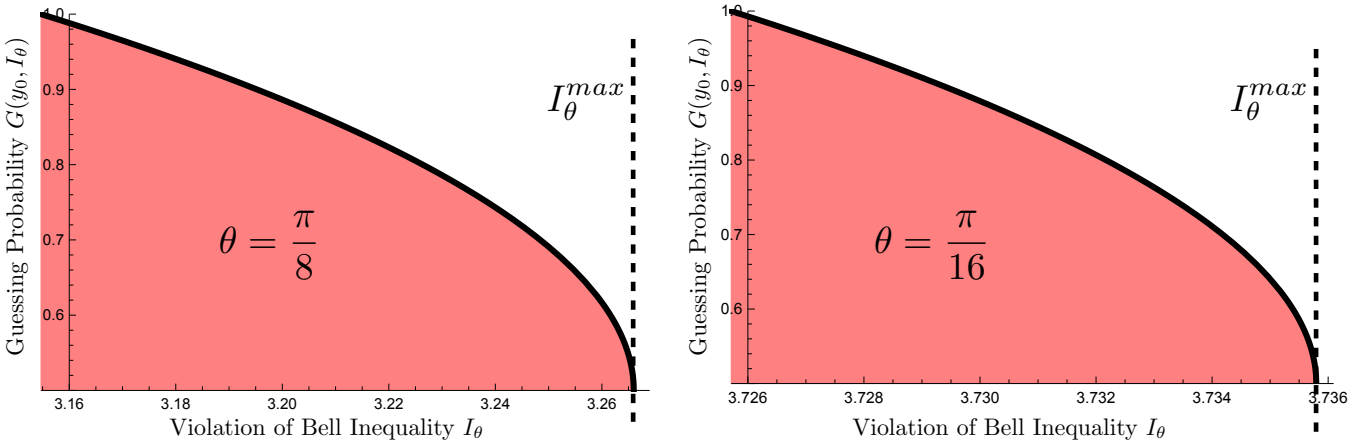

FIG. 3: The upper bound on the guessing probability, this time in function of the violation of $I_{\theta=\frac{\pi}{8}}$ and $I_{\theta=\frac{\pi}{16}}$.