Free randomness amplification using bipartite chain correlations

Andrzej Grudka,[1] Karol Horodecki,[2,3] Michał Horodecki,[2,4] Paweł Horodecki,[2,5]
Marcin Pawłowski,[4,6] and Ravishankar Ramanathan[2,4]
[1]*Faculty of Physics, Adam Mickiewicz University, 61-614, Poznań, Poland*
[2]*National Quantum Information Center of Gdańsk, 81-824, Sopot, Poland*
[3]*Institute of Informatics, University of Gdańsk, 80-952, Gdańsk, Poland*
[4]*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952, Gdańsk, Poland*
[5]*Faculty of Applied Physics and Mathematics, Technical University of Gdańsk, 80-233, Gdańsk, Poland*
[6]*Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK*
(Received 14 February 2014; revised manuscript received 26 August 2014; published 22 September 2014)

A direct analysis of the task of randomness amplification from Santha-Vazirani sources using the violation of the chained Bell inequality is performed in terms of the convex combination of no-signaling boxes required to simulate quantum violation of the inequality. This analysis is used to find the exact threshold value of the initial randomness parameter from which perfect randomness can be extracted in the asymptotic limit of a large number of measurement settings. As a byproduct, we provide a tool for the analysis of randomness amplification protocols, namely a general characterization of the probability distributions of bits generated by Santha-Vazirani sources, which are shown to be mixtures of specific permutations of Bernoulli distributions with a parameter defined by the source.

DOI: 10.1103/PhysRevA.90.032322 PACS number(s): 03.67.Ac, 03.65.Ta, 03.65.Ud

I. INTRODUCTION

The question whether all processes in Nature are predetermined or if there are fundamentally unpredictable events is a most fundamental one. While it seems impossible to rule out complete determinism at all levels, the philosophical and practical implications such as in gambling and cryptographic scenarios have made it a question worthy of thorough investigation. In this regard, exciting new results have been obtained by the authors of [1–5] that the correlations in quantum systems can be used to amplify randomness. In particular, it has been shown that the presence of a small amount of unpredictability can be used to infer the presence of truly random events under certain assumptions about the source of unpredictability.

Formally, the information-theoretic task is called randomness amplification, where the goal is to use an input source of partially random bits to produce a perfect random bit. The source of randomness is taken to be the Santha-Vazirani (SV) source [6], defined by the condition that for any random variable $X = (X_1, X_2, \ldots, X_n)$ produced by this source and for any $0 \leqslant i < n$ and $x_i = \{0,1\}$, there holds

$$\tfrac{1}{2} - \epsilon \leqslant P(X_{i+1} = x_{i+1} | X_i = x_i, \ldots, X_1 = x_1) \leqslant \tfrac{1}{2} + \epsilon. \tag{1}$$

The interpretation is that each bit is obtained by the flip of a biased coin, the bias being fixed by an adversary who has knowledge of the history of the process. As such, the conditioning variables can be any set of pre-existing variables W that could be a possible cause of the succeeding bit X_{i+1}. Each bit produced by the source is ϵ free in the sense that the probability distribution is ϵ away in variational distance from uniform. The goal of randomness amplification is to produce perfect random bits, i.e., with $\epsilon_{\text{new}} = 0$. Note that randomness amplification differs from the task of (device-independent) randomness expansion, where it is assumed that an input seed of perfect random bits is available and the goal is to expand this given bit string into a larger sequence of random bits. Quantum nonlocality has also found application in this later task [7–9] as well as in device-independent cryptographic scenarios [10,11].

In [6], it was shown that the randomness produced by a single SV source cannot be amplified by classical means, by any deterministic function. The idea behind randomness amplification using quantum correlations in [1,2,4,5] is to use the SV source to choose the measurement settings of a set of spatially separated observers in a Bell test and to obtain random bits from some function of the measurement outcomes. In [1], the bipartite scenario of chained Bell inequalities [12] was shown to be useful in obtaining perfectly random bits as measurement outcomes for a limited range of ϵ values ($\epsilon < \frac{(\sqrt{2}-1)^2}{2}$ assuming correctness of quantum theory). The validity of the no-signaling principle is vital in the protocol, no-signaling being necessary for perfect randomness to occur in any theory. While in [1], it was recognized that the chain inequality could not be used to amplify arbitrarily weak randomness, an open problem was to determine the precise range of ϵ from which randomness could be amplified. This is one of the major questions we address in the paper.

Let us also discuss the differences between the protocol in [1] and other protocols designed for this task. In [2], a different protocol was proposed to generate perfect random bits from any initial value of $\epsilon < \frac{1}{2}$. This later protocol differs from [1] in at least three respects. First, it involves the use of the five-party Mermin inequality and a corresponding noiseless Greenberger-Horne-Zeilinger state, and hence leaves an open question whether an analogous result (i.e., amplification of arbitrarily weak randomness) can be derived using a bipartite physical system. Perhaps more crucially, the major drawback of the protocol is that it requires a large number of space-like separated devices for its implementation, with the number of devices growing with the randomness of the final bit obtained and tending to infinity in the limit of a perfectly random output. Finally, the hashing function used to compute the final random

bit is not explicitly provided but only proven to exist. As a result, while the protocol of [2] is of theoretical interest, it is of limited practical value. Hence, a thorough analysis and the determination of the exact limits of the protocol in [1] (which we perform here) is important both from the fundamental perspective as well as from the practical benefits of a two-party protocol in view of the constraint of space-like separation.

A fundamental understanding of the probability distributions of bits generated by the source of partial randomness is also necessary to study how and when tasks such as randomness amplification can be performed given different strengths of the adversary. A central result of this paper is an investigation into the structure of the SV source showing that the extremal points of the set of probability distributions from such a source are permutations of Bernoulli distributions. Indeed, this fact has found application in randomness amplification against adversaries limited to quantum resources [3]. Moreover, in the search for simpler (possibly bipartite) protocols for generating perfect random bits from any initial value of ϵ against no-signaling adversaries, it becomes vital to derive intuitive methods that apply to arbitrary scenarios as well as to understand the limits of applicability of currently known protocols [1]. We address these issues, providing an analysis of the task of amplification in terms of the randomness present in the no-signaling boxes that appear in a convex decomposition of the quantum box of probabilities. This is used to derive the optimal range of ϵ values from which perfect randomness can be generated using the bipartite chain correlations as well as to extend the result to the determination of the exact threshold value in the asymptotic limit of a large number of measurement settings. See also [13] for a different approach to obtaining randomness from the chain inequalities using a trusted third party (a referee).

II. STRUCTURE OF SANTHA-VAZIRANI SOURCES

The protocol for randomness amplification from SV sources using nonlocal correlations involves using the source to choose the measurement settings in the Bell test. In the bipartite chain Bell test, each party uses a string of bits from a source to generate the measurement settings $\mathbf{x}$ and $\mathbf{y}$, the sources held by the parties may be correlated with each other. Our aim in this section is to characterize the probability distributions $Q(\mathbf{x},\mathbf{y}|w)$ which can arise from the source given any other random variable w possibly held by an adversary Eve. We investigate the structure of the SV sources and prove that the distributions obeying (1) are mixtures of permuted Bernoulli distributions. Formally, we state the following proposition (see proof in the Appendix).

Proposition 1. Extremal points of the set of probability distributions from a Santha-Vazirani source with parameter ϵ are permutations of Bernoulli distributions with parameter $p = p_+$, where $p_+ = \frac{1}{2} + \epsilon$.

Note that not all permutations are allowed, in the proof we provide a detailed explanation of the allowed permutations. We apply the characterization from Proposition 1 in the following sections to find the optimal values of ϵ from which randomness can be amplified using the chain correlations.

III. RANDOMNESS AMPLIFICATION FROM NONLOCAL QUANTUM CORRELATIONS

Consider that the bits generated by the SV source (that are partially free with respect to any set of space-time variables held by an adversary Eve) are used to choose the measurement settings in a Bell test by a set of N spatially separated observers. Upon violation of the inequality, the parties process the measurement outcomes to obtain a perfect random bit. The general N party Bell inequality for randomly chosen measurement settings can be written as

$$\beta = \sum_{\vec{a},\vec{x}} \alpha(\vec{a},\vec{x}) P(\vec{a}|\vec{x}) \geqslant \beta_L. \tag{2}$$

Here, $\vec{x}$ denotes a set $\{x_1,\ldots,x_N\}$ of measurement settings chosen by the N parties, $\vec{a} = \{a_1,\ldots,a_N\}$ denotes the outcomes, $\alpha(\vec{a},\vec{x})$ are a set of coefficients and $P(\vec{a}|\vec{x})$ denotes the conditional probability of outcomes $\vec{a}$ given settings $\vec{x}$. The bound β_L denotes the optimal (here, minimal) value of the Bell parameter attainable in local hidden variable (LHV) theories. A quantum state under suitable measurement settings then generates the box B_Q that leads to maximal violation of the inequality β_Q. In the scenario where the measurement settings are not chosen freely but using an ϵ SV source, one obtains a new LHV optimal value as a function of ϵ, $\beta_L(\epsilon)$. The adversary may attempt to simulate the value β_Q using a convex combination of no-signaling boxes $B_{NS}^{(i)}$ which produce values $\beta_{NS}^{(i)}$, i.e., $\beta_Q = \sum_i p_i \beta_{NS}^{(i)}$ with $\sum_i p_i = 1$. The process of randomness amplification is then transparently based on the randomness present in the boxes $B_{NS}^{(i)}$. If some function of the measurement outcomes (in particular, simply one of the outcomes of one party [1]) is random for all boxes $B_{NS}^{(i)}$ appearing in any convex decomposition, then the parties may use this as the output free random bit uncorrelated from Eve. It is immediately seen that to perform free randomness amplification ($\epsilon_{\text{new}} = 0$) from any initial value of $\epsilon < \frac{1}{2}$, one requires that the maximum no-signaling violation of the Bell inequality be achievable within quantum theory; if not, Eve may choose a finite fraction of deterministic boxes in the simulation. In general, for any given β_Q one may write

$$\beta_Q = (1-\delta)\beta_{NS}^{(r)} + \delta\beta_{NS}^{(nr)}, \tag{3}$$

where $\beta_{NS}^{(nr)}$ is the optimal violation of the inequality by boxes without randomness and δ is the maximum fraction of such boxes that Eve may use to simulate β_Q.

IV. RANDOMNESS AMPLIFICATION USING CHAIN INEQUALITIES

In [1], randomness amplification using chained Bell inequalities [12] was investigated. Two results were obtained, the first under the assumption of correctness of quantum theory, i.e., that the observed distribution of measurement outcomes is as given by the theory, and the second without this restriction. In the former scenario, it was shown that for given $\epsilon < \frac{(\sqrt{2}-1)^2}{2}$, there exists a protocol that uses ϵ-free bits with respect to any set of space-time variables W to obtain ϵ'-free bits with respect to W for any $0 \leqslant \epsilon' \leqslant \epsilon$. In particular, the correlations between the outcomes were used to show that the output bit of one

party's measurement is arbitrarily close to being uniform and uncorrelated with W. Following the general considerations of the previous section, we now formulate an intuitive and simpler derivation of this result.

The chained Bell inequality considers the scenario of two spatially separated parties Alice and Bob who each choose from a set of N measurement settings: $x \in \{0,\ldots,N-1\}_A$ for Alice and $y \in \{0,\ldots,N-1\}_B$ for Bob. Each measurement results in a binary outcome $a \in \{0,1\}$ for Alice and $b \in \{0,1\}$ for Bob. The inequality is written as

$$\sum_{x=y \text{ or } x=y+1} \sum_{a,b} P(a \oplus b = 1|x,y) + P(a \oplus b = 0|0,N-1) \geqslant 1, \quad (4)$$

where $\oplus$ denotes addition modulo 2. Notice that out of the N^2 possible measurement pairs, only the $2N$ neighboring pairs where $x = y$ or $x = y+1$ (sum modulo N) forming a chain are considered in the inequality. The LHV bound is obtained from the fact that perfect correlations in the outcomes for $2N-1$ pairs in the sum automatically implies perfect correlation for the pair $\{0,N-1\}$. Quantum mechanics violates this inequality obtaining a value of $2N\sin^2(\frac{\pi}{4N})$, which for large N tends to the algebraic limit of 0. This optimal value is obtained by measuring on the maximally entangled state $|\phi_+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ with the measurement settings defined by the bases $\{|\alpha\rangle, |\alpha+\pi\rangle\}$ for Alice and $\{|\beta\rangle, |\beta+\pi\rangle\}$ for Bob where $|\theta\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle$ and the angles $\alpha \in \frac{\pi}{2N}\{0,2,\ldots,2N-2\}$ and $\beta \in \frac{\pi}{2N}\{1,3,\ldots,2N-1\}$. The set of no-signaling boxes for this scenario was studied in [14], no-signaling boxes with precisely the structure of perfect correlations for the $2N-1$ neighboring pairs in the sum and perfect anticorrelations for the remaining pair exist which in addition to incorporating perfect randomness attain the optimal no-signaling value of 0. A crucial observation is that if one pair of measurement settings in the expression is known to not occur, classical theories can simulate optimal no-signaling violation of the inequality.

Ideally the measurement settings are chosen freely, however, in this scenario they are chosen by Alice and Bob each using $r := \log_2 N$ bits from an SV source with nonzero ϵ. The optimal strategy by an adversary using local boxes is then to choose the term that equals 1 in the Bell expression corresponding to the pair of measurements that the SV source provides with minimum probability. One therefore considers the inequality

$$\sum_{x=y \text{ or } x=y+1} \sum_{a,b} Q(x,y|w)P(a \oplus b = 1|x,y) + Q(0,N-1|w)P(a \oplus b = 0|0,N-1) \geqslant p_{\min} \quad (5)$$

for each w in the set of space-time variables with which the imperfectly free SV source may be correlated (thought of as held by Eve). Here $Q(x,y|w)$ is the probability of a pair of measurement settings x,y being chosen by Alice and Bob and we have the normalization constraint $\sum_{x=y \text{ or } x=y+1} Q(x,y|w) = 1$. The bound $p_{\min} = \min_{x,y} Q(x,y|w)$ is the minimum probability of a pair of measurement settings chosen by Alice and Bob, ideally $p_{\min}^{\text{ideal}} = \frac{1}{2N}$ (for $\epsilon = 0$). As in the previous section, Eve tries to simulate $\beta_Q = \sin^2(\frac{\pi}{4N})$ using no-signaling boxes with randomness which produce $\beta_{NS}^{(r)}$ and those that do not incorporate randomness and give $\beta_{NS}^{(nr)}$. Crucially, for the chained inequalities only those boxes [14] with perfect randomness in the outcomes ($\epsilon_{\text{new}} = 0$) violate the chain inequality giving $\beta_{NS}^{(r)} = 0$. All other no-signaling boxes either do not give randomness or produce $\beta_{NS}^{(nr)} \geqslant p_{\min}$. Therefore, the optimal violation of the inequality that Eve can achieve using any fraction δ of nonrandom no-signaling boxes (and fraction $(1-\delta)$ of random ones) is given by $\beta_{\text{sv}} = \delta p_{\min}$.

The measurement settings are chosen using $2r$ uses of the imperfect SV source, say the first r bits give in binary the setting x for Alice and the next r bits give Bob's setting y. The minimum probability of occurrence for any measurement pair among the N^2 pairs is $p_-^{2r} = (\frac{1}{2}-\epsilon)^{2r}$. From the set of obtained settings, only those corresponding to the $2N$ neighboring pairs in the inequality are retained with the rest discarded. Therefore, the minimum probability in the sequence of $2N$ pairs, $p_{\min}$ is given by

$$p_{\min} = \frac{p_-^{2r}}{p_-^{2r} + ||Q(\mathbf{x},\mathbf{y}|w)||_{2N-1}}, \quad (6)$$

where $||Q(\mathbf{x},\mathbf{y}|w)||_{2N-1}$ is the $(2N-1)^{th}$ Ky Fan norm of the probability distribution $Q(\mathbf{x},\mathbf{y}|w)$ generated by the source, i.e., the sum of the $2N-1$ largest probabilities. The denominator of the above expression is bounded from above by $2Np_+^{2r}$ where $p_+ = (\frac{1}{2}+\epsilon)$ since p_+^{2r} is the largest probability of occurrence of a bit string of length $2r$ from the source. We therefore obtain that the value of the Bell expression simulated by Eve is

$$\beta_{\text{sv}} = \delta p_{\min} \geqslant \delta \frac{p_-^{2r}}{2^{r+1}p_+^{2r}}. \quad (7)$$

For consistency with the quantum value, $\beta_{sv} \leqslant \beta_Q$, i.e., $\delta\frac{p_-^{2r}}{2^{r+1}p_+^{2r}} \leqslant \sin^2(\frac{\pi}{2^{r+2}})$. The fraction of nonrandom boxes δ approaches 0 (perfect randomness is obtained) as we increase the number of settings $N(=2^r)$ provided

$$\lim_{r\to\infty} \frac{\pi^2}{8}\frac{p_+^{2r}}{2^r p_-^{2r}} = 0, \quad (8)$$

giving $\frac{(\frac{1}{2}+\epsilon)^2}{2(\frac{1}{2}-\epsilon)^2} < 1$, thus recovering $\epsilon < \frac{(\sqrt{2}-1)^2}{2} \approx 0.086$.

V. ASYMPTOTICALLY EXACT BOUNDS ON RANDOMNESS

We now show an improved estimate of $p_{\min}$ which gives exact values for the range of allowed ϵ in the asymptotic limit of large N. Among the joint probability distributions that satisfy the SV conditions are the *extremal* ones which as we have seen are (certain) permutations of the Bernoulli distribution. Our goal is to find the $(2N-1)^{th}$ Ky Fan norm of the Bernoulli distribution, which being the sum of the $2N-1$ largest probabilities is permutation invariant and the same for all extremal distributions.

The $2N-1$ Ky Fan norm of the Bernoulli distribution B satisfying (1) is

$$||B||_{2^{r+1}-1} = \sum_{i=0}^{m} \binom{2r}{i} p_+^{2r-i} p_-^i, \tag{9}$$

where m is chosen to obtain the $2^{r+1}-1$ largest probabilities. The task of finding m can be reformulated using

$$m \leqslant \min_c \left\{ cr : \sum_{i=0}^{cr} \binom{2r}{i} \geqslant 2^{r+1}-1 \right\} \tag{10}$$

to finding the minimum c satisfying the inequality above.

We now state the following lemma (proof in the Appendix) to bound $||B||_{2^{r+1}-1}$, leading to the asymptotically exact range of ϵ from which perfect randomness may be extracted.

Lemma 1. The Ky Fan norm of the Bernoulli distribution $||B||_{2^{r+1}-1}$ with parameter $p_- = (1/2-\epsilon)$ for large r obeys

$$\binom{2r}{cr} p_-^{cr} p_+^{(2-c)r} < ||B||_{2^{r+1}-1} < k \binom{2r}{cr} p_-^{cr} p_+^{(2-c)r}, \tag{11}$$

where c is the solution to $2^{2rH(c/2)} = 2^r$ ($c \approx 0.22$) and $k = \frac{(2-c)(1-2\epsilon)}{2(1-c-2\epsilon)}$.

The above lemma is now used to find when the upper bound on δ approaches zero, i.e., when

$$\lim_{r\to\infty} \frac{\pi^2}{16} \frac{||B||_{2^{r+1}-1}}{2^{2r} p_-^{2r}} = 0. \tag{12}$$

The bounds in (11) imply that the limit is defined by the behavior for large r of

$$\frac{\binom{2r}{cr} p_-^{cr} p_+^{(2-c)r}}{2^{2r} p_-^{2r}} \approx 2^{2rH(c/2)} \frac{p_+^{(2-c)r}}{2^{2r} p_-^{(2-c)r}}. \tag{13}$$

For the limit to be 0, we need $(1/2+\epsilon)^{2-c} < 2(1/2-\epsilon)^{2-c}$ for $c \approx 0.22$ giving $\epsilon < \frac{2^{1/(2-c)}-1}{2(2^{1/(2-c)}+1)} \approx 0.0961$. We thus obtain the result that the asymptotically exact maximal value for ϵ is 0.0961. In fact, for ϵ larger than this critical value, free randomness cannot be obtained in the protocol, i.e., $\epsilon_{\text{new}} > \epsilon$ for this range as shown below.

Note that the amount of randomness ϵ_{new} obtained in the protocol is given by $(1/2)+\epsilon_{\text{new}} = (1-\delta)\times(1/2)+\delta\times 1$, i.e., $\epsilon_{\text{new}} = \delta/2$ since for a fraction δ of the boxes we have deterministic outputs and $(1-\delta)$ of the boxes yield perfectly random output. The optimal value of δ is given by $\frac{\beta_Q}{p_{\min}}$. The upper and lower bounds in Lemma 1 converge to the same value in the limit of large N so that we have that $\delta \to 0$ only if $\epsilon < 0.0961$, above this value the violation can be simulated with local boxes. In the scenario where the randomness is certified by a device-independent test rather than by assuming the correlations are as in quantum theory, a similar analysis yields that the optimal value of ϵ below which randomness may be obtained is 0.0623 which again improves the bound in [1].

VI. CONCLUSION

Randomness amplification from quantum nonlocal correlations is shown to be directly related to the fraction of no-signaling boxes incorporating randomness which appear in any possible convex combination of boxes simulating the Bell violation. An intuitive and simple derivation is provided for the range of partial randomness from which perfect randomness can be generated using quantum correlations violating the bipartite chained Bell inequalities. Asymptotically exact bounds on the minimum probability of a pair of measurement settings from an SV source enable us to identify the exact threshold value of the most imperfect source from which perfect randomness can be extracted using these correlations. We note that the results obtained here are *incomparable* to a previous result in [2] which even though it obtains randomness from an arbitrarily weak SV source, requires (i) the use of a multiparty Bell inequality and (ii) a large number of noiseless space-like separated devices to do so (in fact, it requires an infinite number of noiseless no-signaling devices to obtain a perfect random bit). Indeed, it is a major open problem if arbitrarily imperfect random sources can be amplified in a bipartite scenario at all. The present paper contributes to the solution of this open problem by narrowing the region of initial randomness in question. Finally, let us note that the characterization of the probability distributions from the SV source provided here is of independent interest in general scenarios as well [3].

ACKNOWLEDGMENTS

K.H. and M.H. thank Renato Renner for introducing them to the problem of amplification of randomness. The paper is supported by ERC AdG grant QOLAPS, EC grant RAQUEL, and by the Foundation for Polish Science (FNP) TEAM project co-financed by the EU European Regional Development Fund. Part of this work was done in National Quantum Information Center of Gdańsk. K.H. would also like to acknowledge discussion with Justyna Łodyga.

APPENDIX

Here, we present the formal proof of the proposition and lemma stated in the text.

Proposition 1. Extremal points of the set of probability distributions from Santha-Vazirani source are permutations of Bernoulli distributions with parameter $p = p_+$, with $p_+ = \frac{1}{2}+\epsilon$.

To prove the proposition we will need the following lemma.

Lemma 2. Consider two alphabets X and Y, with $|X| = K, |Y| = M$. Consider some convex sets S_X and S_Y of the probability distributions over the spaces X and Y, respectively. Consider an arbitrary joint probability distribution $p(x,y)$. Let $p(y|x)$ be the corresponding conditional probability distribution and $p(x)$ the marginal one. Suppose now that for any fixed x, the distribution $\{p(y|x)\}_y$ belongs to S_Y, and the distribution $\{p(x)\}_x$ belongs to S_X. Then we can write $p(x,y)$ as a mixture of probability distributions of the form

$$\tilde{p}(x,y) = p^{(x)}(y) r(x), \tag{A1}$$

where distribution $p^{(x)}$ is extremal in the set S_Y and distribution r is extremal in set S_X.

Proof. Clearly, it is enough to prove that $p(x,y)$ can be written as mixture of distributions

$$p'(x,y) = p(x) p^{(x)}(y), \tag{A2}$$

where $p^{(x)}$ is extremal in S_Y. Indeed, then we can decompose $p(x)$ into extremal points in S_X, and reach the form (2).

Let $p^{(i)}$ run over extremal elements of S_Y. We define the following distributions

$$p_{i_1,\ldots,i_K}(x,y) = p(x)p^{(i_x)}(y). \tag{A3}$$

Clearly they are of the required form (A2). We will now show that a suitable mixture of such distributions gives $p(x,y)$. To see this, note that since for each x, the distribution $p(y|x)$ belongs to S_Y, we can write it as a mixture of $p^{(i)}$'s

$$p(y|x) = \sum_i \lambda_i^{(x)} p^{(i)}(y), \tag{A4}$$

where

$$\sum_i \lambda_i^{(x)} = 1, \tag{A5}$$

for each x. We will now show that

$$p(x,y) = \sum_{i_1,\ldots,i_N} \lambda_{i_1}^{(1)} \cdots \lambda_{i_N}^{(N)} p_{i_1,\ldots,i_K}(x,y), \tag{A6}$$

which is what we need to prove, as by (A5) we have

$$\sum_{i_1,\ldots,i_N} \lambda_{i_1}^{(1)} \cdots \lambda_{i_N}^{(N)} = 1 \tag{A7}$$

and $p_{i_1,\ldots,i_K}(x,y)$ are of the required form (A2). To prove the equality (A6), we write

$$\begin{aligned}
&\sum_{i_1,\ldots,i_N} \lambda_{i_1}^{(1)} \cdots \lambda_{i_N}^{(N)} p_{i_1,\ldots,i_K}(x,y) \\
&\quad = \sum_{i_1,\ldots,i_N} \lambda_{i_1}^{(1)} \cdots \lambda_{i_N}^{(N)} p(x) p^{(i_x)}(y) \\
&\quad = \sum_{i_x} \lambda_{i_x}^{(x)} p(x) p^{(i_x)}(y) = p(x)p(y|x) = p(x,y).
\end{aligned} \tag{A8}$$

The last but one equality we obtain from the fact that only for index i_x the summand is nontrivial, for other indices the summands are just λ's, which sum up to 1. ■

Now we are in position to prove Proposition 1.

Proof of Proposition 1. To prove the proposition, we will apply the lemma iteratively. The set X will be the set of n bits, while the set Y will correspond to a single bit. S_Y then has two extremal points (p_+,p_-) and (p_-,p_+). Let us first illustrate the lemma for the case of X also being a single bit. Then simply

$$\{p(x,y)\} = (p(0)p(0|0), p(0)p(1|0), p(1)p(0|1), p(1)p(1|1)). \tag{A9}$$

Now, for $x = 0$, we have decomposition

$$\begin{aligned}
p(0|0) &= \alpha_0 p_+ + (1-\alpha_0)p_-, \quad p(1|0) \\
&= \alpha_0 p_- + (1-\alpha_0)p_+.
\end{aligned} \tag{A10}$$

For $x = 1$ we have some other decomposition

$$\begin{aligned}
p(0|1) &= \alpha_1 p_+ + (1-\alpha_1)p_-, \quad p(1|1) \\
&= \alpha_1 p_- + (1-\alpha_1)p_+.
\end{aligned} \tag{A11}$$

To catch up with notation of the lemma, we have $\alpha_0 = \lambda_1^0, 1-\alpha_0 = \lambda_2^0$, and $\alpha_0 = \lambda_1^{(0)}, 1-\alpha_0 = \lambda_2^{(0)}$, and $p^{(1)} = (p_+,p_-), p^{(2)} = (p_-,p_+)$ are extremal points from S_Y. We can directly check that

$$\begin{aligned}
&(p(0)p(0|0), p(0)p(1|0), p(1)p(0|1), p(1)p(1|1)) \\
&\quad = \alpha_0\alpha_1(p(0)p_+, p(0)p_-, p(1)p_+, p(1)p_-) \\
&\qquad + \alpha_0(1-\alpha_1)(p(0)p_+, p(0)p_-, p(1)p_-, p(1)p_+) \\
&\qquad + (1-\alpha_0)\alpha_1(p(0)p_-, p(0)p_+, p(1)p_+, p(1)p_-) \\
&\qquad + (1-\alpha_0)(1-\alpha_1)(p(0)p_-, p(0)p_+, p(1)p_-, p(1)p_+).
\end{aligned} \tag{A12}$$

Thus we have shown explicitly the decomposition of $p(x,y)$ into distributions of the form (A2). Now we further decompose the distribution $(p(0),p(1))$ into extremal points of S_X which are in this case the same as those of S_Y: (p_+,p_-) and (p_-,p_+). Therefore $\{p(x,y)\}$ is a mixture of the eight probability distributions

$$\begin{aligned}
&(p_+p_+, p_+p_-, p_-p_+, p_-p_-), \quad (p_+p_+, p_+p_-, p_-p_-, p_-p_+), \\
&(p_+p_-, p_+p_+, p_-p_+, p_-p_-), \quad (p_+p_-, p_+p_+, p_-p_-, p_-p_+), \\
&(p_-p_+, p_-p_-, p_+p_+, p_+p_-), \quad (p_-p_+, p_-p_-, p_+p_-, p_+p_+), \\
&(p_-p_-, p_-p_+, p_+p_+, p_+p_-), \quad (p_-p_-, p_-p_+, p_+p_-, p_+p_+),
\end{aligned} \tag{A13}$$

where the ordering is as follows:

$$(p(0,0), p(0,1), p(1,0), p(1,1)). \tag{A14}$$

Note that the first distribution is precisely the Bernoulli distribution, with the probability of 0 in a single trial being $p = p_+$. This distribution is memoryless. The other distributions are not memoryless, but are related to the Bernoulli distribution by permutation of probabilities (not bits). Note that only 8 out of 24 permutations appear.

For n bits, the lemma implies that the extremal probability distributions are created from the product of the extremal distributions for $n-1$ bits, as follows. For a given extremal distribution $(r(1),\ldots,r(K))$ with $K = 2^{n-1}$, we construct the following extremal point:

$$(r(1)p_+, r(1)p_-, r(2)p_+, r(2)p_-, \ldots, r(K)p_+, r(K)p_-). \tag{A15}$$

The other extremal points can be generated from it by changing the order of p_+ and p_- for each $x = 1,\ldots,K$ independently. This implies, that all the extremal points are permutations of the above one. Now, by induction we assume that the distribution $[r(1),\ldots,r(K)]$ over $n-1$ bits is a permutation of Bernoulli distribution over n bits with parameter $p = p_+$. Thus, there is a permutation σ that reorders it, so that it becomes Bernoulli. We can apply this permutation to reorder pairs $[r(i)p_+, r(i)p_-]$ in the distribution (A15). The resulting distribution is Bernoulli for n bits with parameter p_+. Thus (A15) is a permutation of Bernoulli, and hence all other extremal points are too since they are its permutations. Note that not all permutations are allowed because the above construction has the structure of a tree. ■

Lemma 1. The Ky Fan norm of the Bernoulli distribution $||B||_{2^{r+1}-1}$ with parameter $p_- = (1/2-\epsilon)$ is bounded for large

r by

$$\binom{2r}{cr} p_-^{cr} p_+^{(2-c)r} < ||B||_{2^{r+1}-1} < k \binom{2r}{cr} p_-^{cr} p_+^{(2-c)r}, \quad \text{(A16)}$$

where c is the solution to $2^{2rH(c/2)} = 2^r$ [$H(x)$ denotes binary entropy giving $c \approx 0.22$] and $k = \frac{(2-c)(1-2\epsilon)}{2(1-c-2\epsilon)} \approx \frac{0.89(1-2\epsilon)}{2(0.39-\epsilon)}$.

Proof. As seen in the text, the Ky Fan norm of the Bernoulli distribution

$$||B||_{2^{r+1}-1} = \sum_{i=0}^{m} \binom{2r}{i} p_+^{2r-i} p_-^i \quad \text{(A17)}$$

can be reformulated using

$$m \leqslant \min_c \left\{ cr : \sum_{i=0}^{cr} \binom{2r}{i} \geqslant 2N-1 \right\} \quad \text{(A18)}$$

into finding the minimum c that satisfies the inequality above. Note that for $c < 1$

$$\binom{2r}{cr} < \sum_{i=0}^{cr} \binom{2r}{i} < (cr+1)\binom{2r}{cr}, \quad \text{(A19)}$$

since $\binom{2r}{cr}$ is the largest term in the sum. For large r (and consequently large $N = 2^r$), by the Stirling approximation, we have that $\binom{2r}{cr} \approx 2^{2rH(c/2)}$ where $H(x)$ denotes the binary entropy. Therefore, from $\sum_{i=0}^{cr}\binom{2r}{i} \geqslant 2N-1 (\approx 2^r)$ we obtain the condition

$$2^{2rH(c/2)} = 2^r \quad \text{(A20)}$$

giving the value $c \approx 0.22$, which is asymptotically exact because of the inequalities in (A19).

Note that then $||B||_{2^{r+1}-1}$ is trivially lower bounded by $\binom{2r}{cr} p_-^{cr} p_+^{(2-c)r}$ as these form a subset of the probabilities appearing in $||B||_{2^{r+1}-1}$. To derive the upper bound, we use the observation that for $0 \leqslant i \leqslant cr$

$$\frac{\binom{2r}{i-1} p_-^{i-1} p_+^{2r-i+1}}{\binom{2r}{i} p_-^i p_+^{2r-i}} < \frac{i}{2r-i}\frac{p_+}{p_-} \leqslant \alpha, \quad \text{(A21)}$$

where the constant $\alpha = \frac{c(1+2\epsilon)}{(2-c)(1-2\epsilon)} < 1$ for $\epsilon < 0.39$. Iteratively applying the inequality, for $0 \leqslant i \leqslant cr$

$$\binom{2r}{i} p_-^i p_+^{2r-i} < \alpha^{cr-i} \binom{2r}{cr} p_-^{cr} p_+^{(2-c)r}. \quad \text{(A22)}$$

Consequently, we obtain for the Ky Fan norm

$$\begin{aligned} ||B||_{2^{r+1}-1} &< \sum_{i=0}^{cr} \alpha^{cr-i} \binom{2r}{cr} p_-^{cr} p_+^{(2-c)r} \\ &< \binom{2r}{cr} p_-^{cr} p_+^{(2-c)r} \sum_{i=0}^{\infty} \alpha^i \\ &< \frac{(2-c)(1-2\epsilon)}{2(1-c-2\epsilon)} \binom{2r}{cr} p_-^{cr} p_+^{(2-c)r}, \end{aligned} \quad \text{(A23)}$$

which establishes the upper bound. ■

[1] R. Colbeck and R. Renner, Nat. Phys. **8**, 450 (2012).
[2] R. Gallego, L. Masanes, G. de la Torre, C. Dhara, L. Aolita, and A. Acin, Nat. Commun. **4**, 2654 (2013).
[3] P. Mironowicz, R. Gallego and M. Pawłowski, arXiv:1301.7722.
[4] F. G. S. L. Brandao, R. Ramanathan, A. Grudka, K. Horodecki, M. Horodecki, and P. Horodecki, arXiv:1310.4544.
[5] R. Ramanathan, F. G. S. L. Brandao, A. Grudka, K. Horodecki, M. Horodecki, and P. Horodecki, arXiv:1308.4635.
[6] M. Santha and U. V. Vazirani, in *Proceedings of the 25th IEEE Symposium on Foundations of Computer Science, FOCS-84* (IEEE, New York, 1984), p. 434.
[7] S. Pironio *et al.*, Nature (London) **464**, 1021 (2010).
[8] R. Colbeck, Ph.D. dissertation, University of Cambridge, 2007.
[9] A. Acin, S. Massar, and S. Pironio, Phys. Rev. Lett. **108**, 100402 (2012).
[10] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[11] E. Hänggi, R. Renner, and S. Wolf, Eurocrypt **2010**, 216 (2010), arXiv:0911.4171.
[12] S. L. Braunstein and C. M. Caves, Ann. Phys. (NY) **202**, 22 (1990).
[13] R. Augusiak, M. Demianowicz, M. Pawlowski, J. Tura, and A. Acin, arXiv:1307.6390.
[14] N. S. Jones and L. Masanes, Phys. Rev. A **72**, 052312 (2005).