Probing the quantum-classical boundary with compression software

Hou Shun Poh[1], **Marcin Markiewicz**[2], **Paweł Kurzyński**[1,3], **Alessandro Cerè**[1], **Dagomir Kaszlikowski**[1,4], and **Christian Kurtsiefer**[1,4]

[1]Center for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543

[2]Institute of Physics, Jagiellonian University, ul. Stanislawa Lojasiewicza 11, 30-348 Krakow, Poland

[3]Faculty of Physics, Adam Mickiewicz University, Umultowska 85, 61-614 Poznań, Poland

[4]Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542

Abstract. We adapt an algorithmic approach to the problem of local realism in a bipartite scenario. We assume that local outcomes are simulated by spatially separated universal Turing machines. The outcomes are calculated from inputs encoding information about a local measurement setting and a description of the bipartite system sent to both parties. In general, such a description can encode some additional information not available in quantum theory, i.e., local hidden variables. Using the Kolmogorov complexity of local outcomes we derive an inequality that must be obeyed by any local realistic theory. Since the Kolmogorov complexity is in general uncomputable, we show that this inequality can be expressed in terms of lossless compression of the data generated in such experiments and that quantum mechanics violates it. Finally, we confirm experimentally our findings using pairs of photons polarisation-entangled and readily available compression software. We argue that our approach relaxes the i.i.d. assumption, namely that individual bits in the outcome bit-strings do not have to be *independent and identically distributed*.

PACS numbers: 03.67.-a, 03.65.Ta, 42.50.Dv, 89.20.Ff

1. Introduction

In a standard Bell scenario [1] Alice and Bob share a bipartite system and each of them performs a randomly chosen local measurement on their subsystems. Next, Alice and Bob evaluate correlations between their outcomes. A violation of a suitable correlation-based Bell inequality refutes local realism.

The correlations are obtained by repeating the measurements on independent and identically distributed (i.i.d.) pairs and estimating the statistical frequencies $p(x, y|a, b) = N(x, y|a, b)/N(a, b)$. $N(x, y|a, b)$ is the number of times outcomes x and y are obtained when the measurements settings are a and b, whereas $N(a, b)$ is the total number of measurements with settings a and b.

An interesting information theoretic approach to Bell inequalities was proposed in the 80's by Braunstein and Caves [2]. They constructed a test of local realism using conditional Shannon entropy $H(a|b) = H(ab) - H(b)$, where $H(x) = -\sum_i p(x = x_i) \log_2 p(x = x_i)$. This is a measure of how much information about Alice's outcomes is contained in Bob's ones. Although these inequalities are not tight, their immediate advantage is that one does not need to worry about labeling, and they work for any number of outcomes.

The simplest information-theoretic Bell inequality is

$$H(a_1|b_1) \leq H(a_1|b_1) + H(b_1|a_2) + H(a_2|b_1). \tag{1}$$

It holds for local realism but it is violated by quantum theory.

Although the method of Braunstein and Caves offers a conceptually new approach, experimentally it requires the estimation of probabilities $p(x, y|a, b)$. Therefore, actual implementations of such information-theoretic tests are akin to the standard ones and require a similar statistical analysis of the data strings obtained by Alice and Bob's measurements outcome.

Shannon entropy of a data string generated by an i.i.d. source has an important operational meaning. It tells us how much such a data string can be losslessly compressed [3]. Still, Shannon's source coding theorem is based on infinite data strings. In realistic situations data strings are finite and one faces a problem of finding a suitable algorithm for an efficient compression of a given finite data string.

For most data strings it is hard, or impossible, to prove that a compression algorithm we found is the best. However, this does not stop us from introducing a concept of the best possible compression algorithm for a given data string x. This is exactly the idea behind the Kolmogorov complexity. More formally, the concept of Kolmogorov complexity requires a reference to a universal model of computation, for example a universal Turing machine (UTM). In this case the Kolmogorov complexity $K(x)$ of a data string x is the length $l(\Lambda)$ of the shortest program Λ, which, when fed into a UTM, produces an output string x. $K(x)$ is in general uncomputable, however realistic compression algorithms $C(x)$ bound it from above [4], i.e., $K(x) \leq C(x)$. In addition, compression algorithms can be applied to data strings that are generated by non i.i.d. sources.

We show that one can observe violations of local realism by studying compression rates $C(x)$ of realistic compression algorithms applied to outcomes of Bell tests. We derive a compression-based Bell inequality. Next, we experimentally test our inequality using a source of pairs of entangled photons. We observe a violation for properly chosen local measurement settings and a properly chosen compression algorithm. We note that our approach is related to an earlier one by Fuchs [5] and that an alternative approach to non i.i.d. sources was discussed by Gill [6].

2. Algorithmic approach to Bell scenario

Let us start with the following assumption: *If there exists a local- realistic description of Bell measurements, it can be encoded as a program for UTMs with output bit strings corresponding to measurement outcomes of a realistic experiment.* This is equivalent to assuming the physical Church-Turing thesis and applying it to hypothetical local hidden variables. The description program would contain the complete information about the physical system that generates the outcomes, provided that a description of the measurement settings is also provided to the UTM (see figure 1 top).

In this work, we consider two spatially separated UTMs: UTM_A and UTM_B. Each UTM is fed a program encoding information about a bipartite system shared between Alice and Bob, and with an additional program encoding a local measurement setting a_j and b_k ($j, k = 0, 1$). Without loss of generality we can assume that both programs are concatenated in a single bit string.

between them result only from the shared information about the preparation (Λ). In addition, UTM_A has no access to a description of a measurement setting b_k that was fed to UTM_B and vice versa. Therefore, the output strings x_j and y_k, which simulate measurement outcomes, are solely computed from inputs $(\Lambda; a_j)$ and $(\Lambda; b_k)$, respectively. Since both UTMs do not communicate, correlations between x_j and y_k can only originate from Λ. This implies that any possible hidden variable must be encoded in Λ.

The simulation follows these steps (see figure 1 bottom). A copy of the program Λ is sent to each party, Alice and Bob. Each party also receives the description and sequence of the measurements, and, after combining it with Λ, feeds it to the local UTM, generating the output strings x and y. Selecting the bits based on the sequence of measurements, we obtain four pairs of bit strings: $\{x_0, y_0\}$, $\{x_0, y_1\}$, $\{x_1, y_0\}$ and $\{x_1, y_1\}$.

2.1. Analysis of the outcome strings

The core of the Bell test is to find a method to acquire some information about possible hidden variables from the measurement outcomes. In our case we have the output strings $\{x_j, y_k\}$, generated either by a computer simulation or by physical systems. We want to infer from them something about the program capable of generating these strings

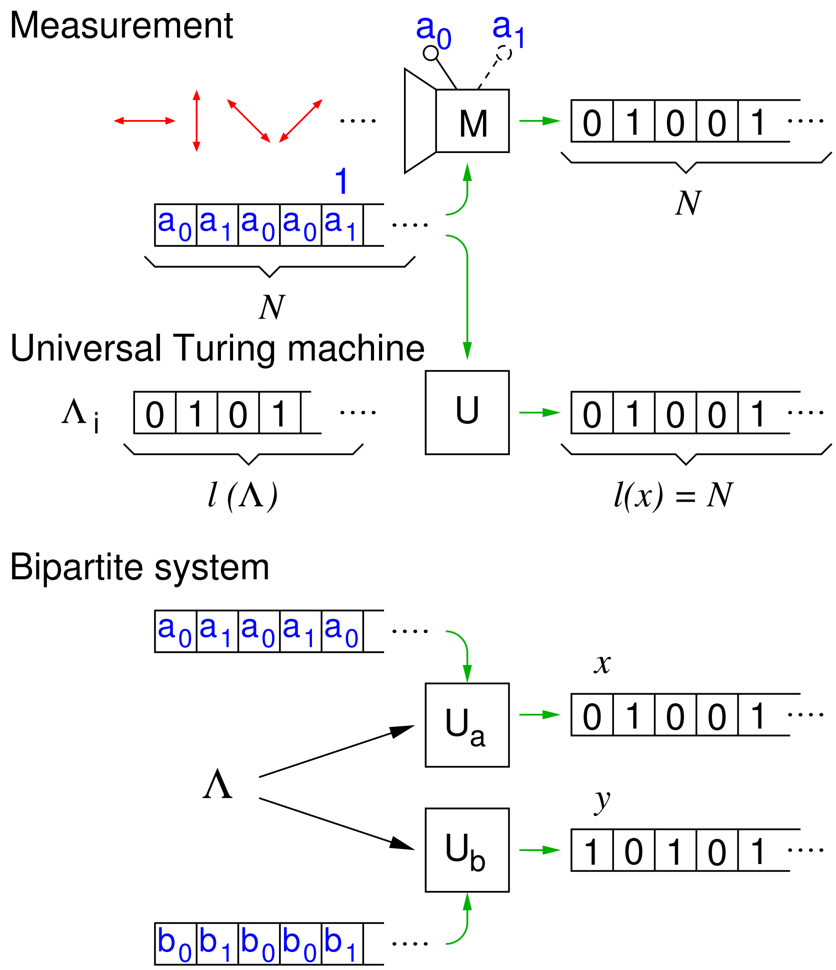

Figure 1. Measurement: N particles enter a measuring device characterized by two polarizer settings a_0 and a_1 generating N-outcome bit strings. A Universal Turing machine (UTM) fed with a program Λ_i and information about the settings a_0 and a_1 can reproduce the string of length N. The bottom part shows a model to reproduce correlated strings x and y generated from measurements on a bipartite system with local UTMs and a common program Λ.

on two spatially separated UTMs. The general idea is to compare the two bit strings using mathematical tools. In the standard approach [1], these tools are statistical, namely one estimates a probability that $x_j^{(n)}$ (the n-th bit of x_j) has the same value as $y_k^{(n)}$. These probabilities can be later used to calculate correlation functions $\langle a_j b_k \rangle$ or entropies $H(a_j|b_k)$, which can be plugged into relevant Bell inequalities. Here, we propose an algorithmic method of comparing these bit strings.

Let us come back to the concept of the Kolmogorov complexity introduced in the beginning. We defined $K(x)$ as the length of the shortest program generating x on a UTM. In the same way we can define $K(x, y)$ as the length of the shortest program generating both strings x and y. If x and y are correlated, $K(x, y) < K(x) + K(y)$, i.e., a UTM can generate y using some parts of the program that was designed for x. In our case we are interested in $K(x_j, y_k)$, $K(x_j)$, and $K(y_k)$. In other words, we seek the shortest hidden variable descriptions of the Bell scenario. Note that if we assume that some hidden variable description exists, then there must exist the shortest such description.

To this end we use the *Normalized Information Distance* (NID) introduced in [7]

$$\mathrm{NID(x,y)} \;=\; \frac{\mathrm{K(x,y)} \;-\; \min\{\mathrm{K(x),K(y)}\}}{\max\{\mathrm{K(x),K(y)}\}}\,. \tag{2}$$

This distance is a metric, up to a correction $\frac{\log_2 N}{N}$, where N is the length of strings x and y. It compares x and y without detailed knowledge about their origin. If the two data strings are identical then $K(x,y) = K(x,x) = K(x)$. The last equality follows from the fact that in order to generate two copies of x we can run the same program twice. In this case $NID(x,y) = 0$. However, if x and y are uncorrelated then $K(x,y) = K(x) + K(y)$ and $NID(x,y) = 1$. In general $0 \leq NID(x,y) \leq 1$.

2.2. Algorithmic Bell inequality

We can use the metric properties of NID to construct a Bell inequality. A similar metric approach to Bell inequalities was used before in [8, 9, 10, 11]. The NID obeys the triangle inequality

$$\mathrm{NID(x,y)} \;+\; \mathrm{NID(y,z)} \;\geq\; \mathrm{NID(x,z)}\,. \tag{3}$$

In our case we can write the following

$$\mathrm{NID(x_0,y_0)} + \mathrm{NID(y_0,y_1)} \geq \mathrm{NID(x_0,y_1)}\,. \tag{4}$$

However, $\mathrm{NID(y_0,y_1)}$ cannot be determined experimentally because the strings y_0 and y_1 cannot be obtained at the same time since they come from incompatible measurements. Thus we follow a standard reasoning used in the derivations of all known Bell inequalities (with a possible exception in [12]). This is called *counterfactual definiteness* and it says that it is admissible to consider outcomes of unperformed experiments. We apply it to our case and assume that uncomputed strings have a definite Kolmogorov complexity. Using the triangle inequality $\mathrm{NID(x_1,y_0)} + \mathrm{NID(x_1,y_1)} \geq \mathrm{NID(y_0,y_1)}$, and combining it with (4) we get:

$$\mathrm{NID(x_0,y_0)} + \mathrm{NID(x_1,y_0)} + \mathrm{NID(x_1,y_1)} \geq \mathrm{NID(x_0,y_1)}\,. \tag{5}$$

As mentioned above, NID is only approximately a metric, therefore the above inequality holds up to a term $\frac{\log_2 N}{N}$.

For convenience we introduce the parameter S' quantifying the degree of violation of inequality (5):

$$S' = \mathrm{NID(x_0,y_1)} - \mathrm{NID(x_0,y_0)} - \mathrm{NID(x_1,y_0)} \;-\; \mathrm{NID(x_1,y_1)} \leq 0 \tag{6}$$

A violation of the local realism hypothesis occurs if S' is positive. To test the positivity of S' we have to address a problem that also appears in standard Bell scenarios (for a suggested resolution see [12]). For example, we can set up a source to generate entangled photon pairs in an arbitrary state. However, for every experimental run i, with the same preparation, the generated local string x_{j_i} can be different. Consequently, the corresponding program Λ_i can be different for every experimental run.

To address this issue, we introduce a new assumption: for every two experimental runs i and i' the complexity of the generated strings remains the same: $K(x_{i,a_j}) = K(x_{i',a_j})$ and $K(x_{i,a_j}, y_{i,b_k}) = K(x_{i',a_j}, y_{i',b_k})$ up to a term $\frac{\log_2 N}{N}$. We call this assumption *uniform complexity*. Thus our inequality only applies to programs that have this property. It follows that if the inequality (6) is violated, either local realism or uniform complexity are invalid. Uniform complexity can in principle be tested experimentally.

In general, Kolmogorov complexity cannot be evaluated, but it can be estimated. One can adapt two conceptually different approaches. The first one, the statistical approach, results in an inequality that is similar to Eq. (1). We will briefly discuss it, however we will focus on the algorithmic one

2.2.1. Statistical approach The statistical approach uses an ensemble X of all possible 2^N bit strings of length N and looks for their average Kolmogorov complexity. The ensemble average is the Shannon entropy $H(X)$ [13] and thus

$$\langle \mathrm{NID(x,y)} \rangle = \frac{\mathrm{H(x,y) - \min\{H(x), H(y)\}}}{\mathrm{\max\{H(x), H(y)\}}} \,. \tag{7}$$

Inequality (5) becomes an entropic Bell inequality [2] if local entropies are maximal, i.e., $H(x) = H(y) = N$. Braunstein and Caves showed that for a maximally entangled polarization state of two photons and polarizer angles such that $\vec{a_0} \cdot \vec{b_1} = \cos 3\theta$, $\vec{a_0} \cdot \vec{b_0} = \vec{a_1} \cdot \vec{b_0} = \vec{a_1} \cdot \vec{b_1} = \cos\theta$, inequality (5) is violated for an appropriate range of θ. An expected quantum value of S' as a function of θ is shown in figure 2a. The maximal violation is $S' = 0.24$ for $\theta = 8.6°$.

2.2.2. Algorithmic approach On the other hand, it is possible to avoid a statistical description of our experiment following Ref. [4] where it was shown that the Kolmogorov complexity can be well approximated by compression algorithms. In [4] the *Normalized Compression Distance* (NCD) is introduced

$$\mathrm{NCD(x,y)} = \frac{\mathrm{C(x,y) - \min\{C(x), C(y)\}}}{\mathrm{\max\{C(x), C(y)\}}} \,, \tag{8}$$

where $C(x)$ is the length of the compressed string x, and $C(x, y)$ is the length of the compressed concatenated strings x, y. Replacing NID with NCD in inequality (6) leads to a new inequality:

$$\begin{aligned} S' \to S = &\, \mathrm{NCD(x_{a_0}, y_{b_1}) - NCD(x_{a_0}, y_{b_0})} \\ &- \mathrm{NCD(x_{a_1}, y_{b_0}) - NCD(x_{a_1}, y_{b_1})} \leq 0 \,. \end{aligned} \tag{9}$$

This expression can be tested experimentally because the NCD is operationally defined. Moreover, it was shown in [4] that NCD is a metric up to a term $\frac{\log_2 N}{N}$, therefore inequality (9) holds up to the same term.

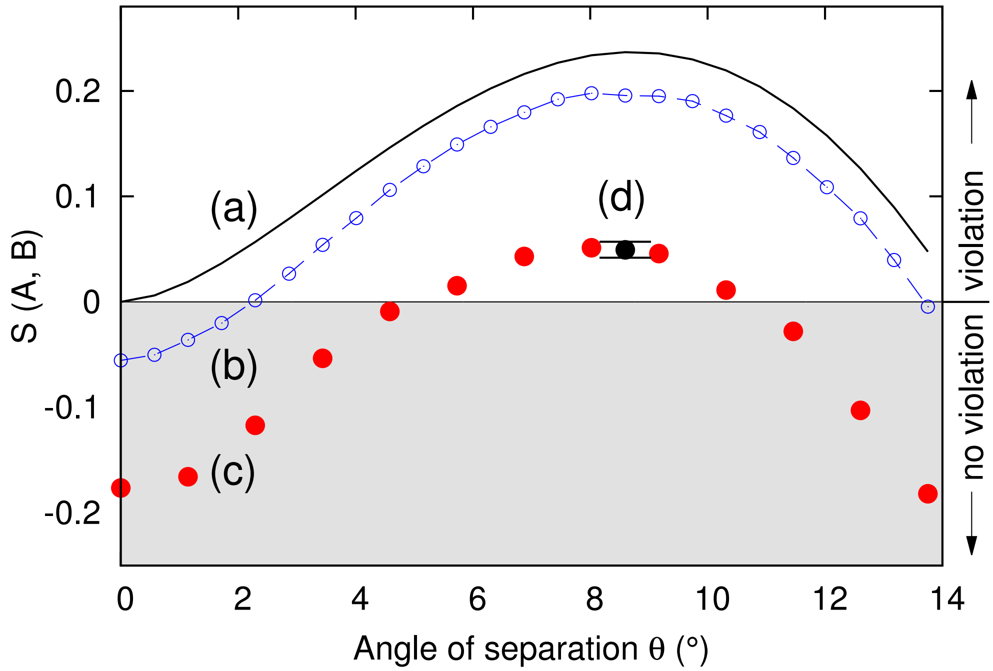

Figure 2. Plots of S versus angle of separation θ. (a) Result obtained from Eq. (7), (b) result obtained from using the LZMA compressor on numerically generated data, (c) measurement of S in the experiment shown in figure 5, and (d) longer measurement at the optimal angle$\theta = 8.6°$.

3. Choice of compressor

Most compression algorithms use some prediction about the data composition. If it matches this prediction, the compression can be done efficiently. To conduct an experiment we need to ensure the suitability of the compression software we use to evaluate the NCD. For this, we numerically simulate the outcome of an experiment based on a distribution of results predicted by quantum physics.

In order to evaluate the NCDs of the binary strings, we need to choose a compression algorithm that performs close to the Shannon limit [3]. This is necessary to ensure that it does not introduce any unintended artifacts that lead to an overestimation of the violation. Preferably we want to work in the regime where the obtained NCDs always underestimate the violation. For this purpose, we characterized four compression algorithms implemented by freely available compression programs: LZMA [14], BZIP2 [15], GZIP [16] and LZW [17]. To eliminate the overhead associated with the compression of ASCII text files, we save data in binary format.

For this characterization, and the following simulation of the experiment, we need to generate a “random” string of bits (1, 0) or pairs of bits (00, 01, 10, and 11) of various length with various probability distributions. We generate these strings using the *MATLAB* [18] function *randsample()* that uses the pseudo random number generator *mt19937ar* with a long period of $2^{19937} - 1$. It is based on the Mersenne Twister [19],

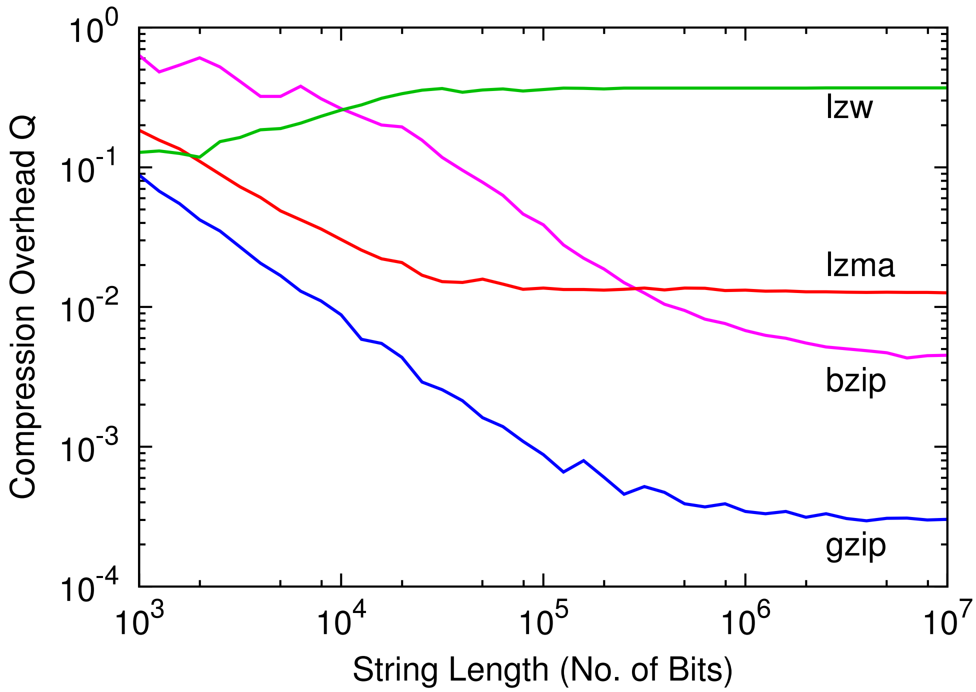

Figure 3. Comparison of the compression overhead Q obtained using four different compression algorithms on pseudo- random strings of varying lengths. The expected value for an ideal compressor is 0. From this characterization we can exclude LZW as a useful compressor for our application.

with ziggurat [20] as the algorithm that generates the required probability distribution. The complexity of this (deterministic) source of pseudorandom numbers should be high enough to *not* be captured as algorithmic.

The first part of this characterization involves establishing the minimum string length required for the compression algorithms to perform consistently. We start by generating binary strings, x, with equal probability of 1's and 0's, i.e. random strings, of varying length. For each x, we evaluate the compression overhead Q as

$$Q = \frac{C(x) - H(x)}{l(x)} \,. \tag{10}$$

For a good compressor, we expect Q to be close to 0. From figure 3, it can be seen that for all the compressors, Q starts to converge after about 10^5 bits, setting the minimum string length required for the compressors to work consistently.

In the second part of this characterization, test the compressors with strings with a known amount of correlation. We generate a random string x of length 10^7 using the same technique already described. We then generate a second string y of equal length and with probability p of being correlated to x. For $p = 0$ the two strings are equal, i.e. perfectly correlated. For $p = 0.5$ they are uncorrelated. Strings x and y are combined to form the string xy. To avoid artifacts due to the limited data block size of the compression algorithms, the elements of x and y are interleaved: for example, for string $x = (0, 0, 0)$ and $y = (1, 1, 1)$, the resulting strings is $xy = (0, 1, 0, 1, 0, 1)$. The same

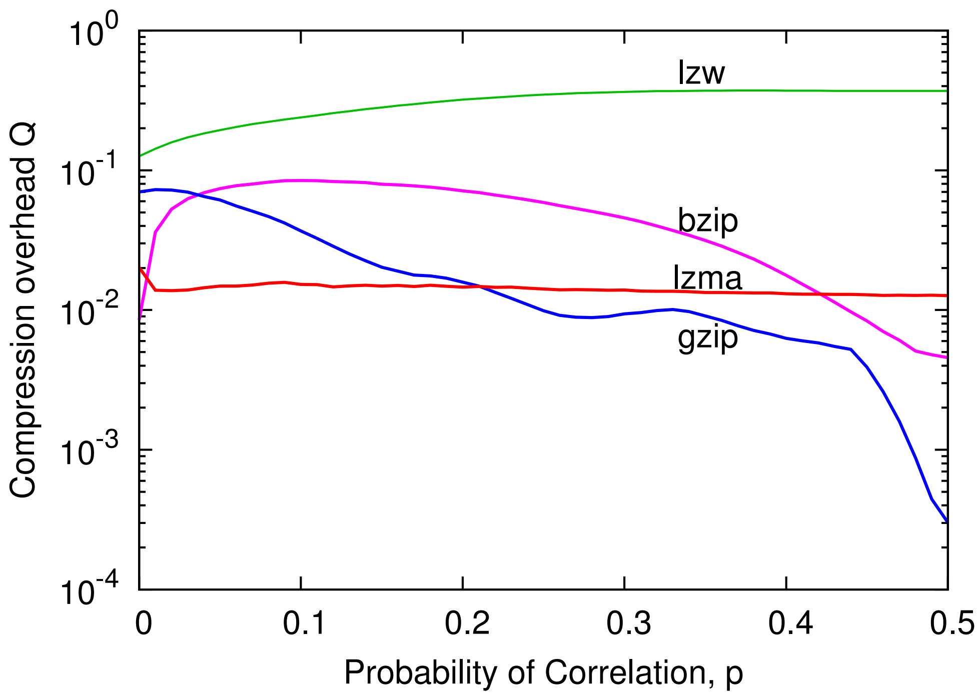

Figure 4. Compression overhead Q for the string xy as a function of the probability of pairwise correlation p between the bits of the generating strings x and y for three different compressors: BZIP2, GZIP, and LZMA.

interleaving procedure is also implemented for the strings generated in the experiment, as described later on. We then compress xy and evaluate the compression overhead Q as a function of p. The results for different compressors are shown in figure 4. Although there are ranges of p where BZIP2 and GZIP perform better than LZMA, the latter shows a more uniform performance over the entire interval of p. On the other hand, LZW performs poorly in all respects. It is reasonable to assume that the use of LZMA should reduce the possibility of artifacts in the estimation of the NCD also for the data obtained from the experiment.

In general our method can be used for data from any source by finding a suitable compression algorithm [4]. Thus, we are not limited to i.i.d. sources, as it is commonly assumed in standard statistical ensemble-based experiments, like, for instance, Bell-type tests.

The numerical simulation also verifies the angle that maximizes the violation of inequality (9). The results of this simulation are presented in figure 2.

4. Experiment

In our experiment (see figure 5), the output of a grating-stabilized laser diode (LD, central wavelength 405 nm) passes through a single mode optical fiber (SMF) for spatial mode filtering, and is focused to a beam waist of 80 μm into a 2 mm thick BBO crystal cut for type-II phase- matching. There, photon pairs are generated via spontaneous

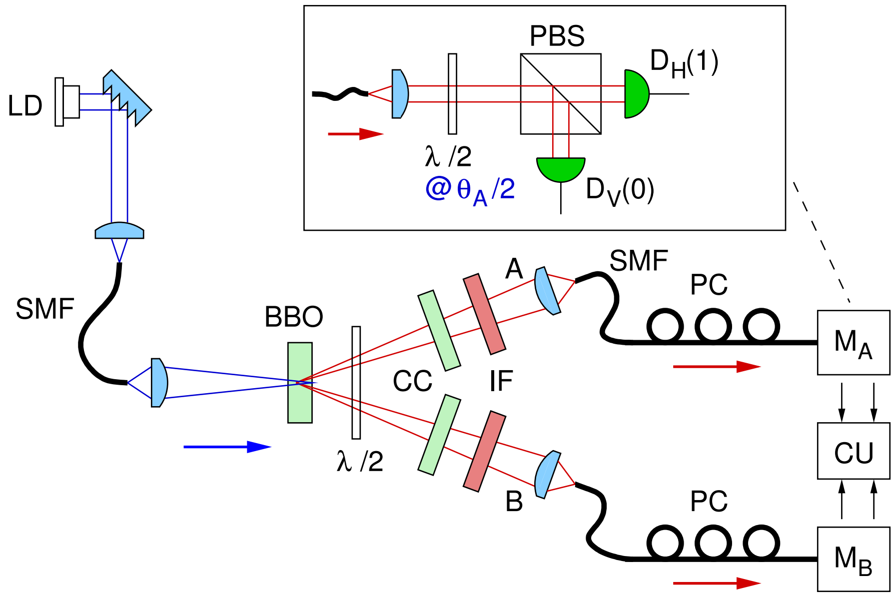

Figure 5. Schematic of the experimental set-up. Polarization correlations of entangled-photon pairs are measured by the polarization analyzers M_A and M_B, each consisting of a half wave plate ($\lambda/2$) followed by a polarization beam splitter (PBS). All photons are detected by Avalanche photodetectors D_H and D_V, and registered in a coincidence unit (CU).

parametric down- conversion (SPDC) in a slightly non-collinear configuration. A half-wave plate ($\lambda/2$) and a pair of compensation crystals (CC) take care of the temporal and transversal walk-off [21]. Two spatial modes (A, B) of down- converted light, defined by the SMFs for 810 nm, are matched to the pump mode to optimize the collection [22]. In type-II SPDC, each down- converted pair consists of an ordinary and extraordinarily polarized photon, corresponding to horizontal (H) and vertical (V) in our setup. A pair of polarization controllers (PC) ensures that the SMFs do not affect the polarization of the collected photons. To arrive at an approximate singlet Bell state, the phase ϕ between the two decay possibilities in the polarization state $|\psi\rangle = 1/\sqrt{2}\left(|H\rangle_A|V\rangle_B + e^{i\phi}|V\rangle_A|H\rangle_B\right)$ is adjusted to $\phi = \pi$ by tilting the CC.

In the polarization analyzers (inset of figure 5), photons from SPDC are projected onto arbitrary linear polarization by $\lambda/2$ plates, set to half of the analyzing angles $\theta_{A(B)}$, and polarization beam splitter (extinction ratio 1/2000 and 1/200 respectively for transmitted and reflected arm) in each analyzer. Photons are detected by avalanche photo diodes (APD), and corresponding detection events from the same pair identified by a coincidence unit if they arrive within $\approx \pm 3$ ns of each other.

The quality of polarization entanglement is tested by probing the polarization correlations in a basis complementary to the intrinsic HV basis of the crystal. With interference filters (IF) of 5 nm bandwidth (FWHM) centered at 810 nm, in the 45° linear polarization basis we observe a visibility $V_{45} = 99.9 \pm 0.1\%$. The visibility in the natural H/V basis of the type-II down- conversion process also reaches $V_{\mathrm{HV}} = 99.9 \pm 0.1\%$. A separate test of a CHSH-type Bell inequality [23] leads to a value of $S = 2.826 \pm 0.0015$. This indicates a high quality of polarization entanglement; the uncertainties in the visibilities are obtained from propagated Poissonian counting statistics.

4.1. Measurement and Data Post-processing

In the realization of this proof of principle experiment, we did not intend to provide a loophole-free demonstration. Due to the limited efficiency of the APD detectors, we assume a fair sampling. Similarly, even if Alice and Bob are not space-like separated, we assume that no communication happens between the two measurements. Moreover, the basis choice is not random, as expected in an ideal Bell-like experiment. We instead set the basis and record the number of events in a fixed time. We are assuming that the state generated by the source, and all the other parameters of the experiment, do not change between experimental runs.

The basic measurement lasts 60 s, during which we record an average of 16×10^3 two-fold coincidences between detectors at A and B. A detection event at the transmitted output of each PBS is associated with 0, reflected one with 1. Three- and four-fold coincidences, as well as two-fold coincidences between detectors belonging to the same party, correspond to multiple pairs of photos generated withing the coincidence time window. The rate of these events is negligible, therefore we discarded them.

In order to avoid biases due to the asymmetries in detector efficiencies, to measure one basis (a_j, b_k) we also measure three complementary basis: $(a_j{+}45°, b_k)$, $(a_j, b_k{+}45°)$, and $(a_j + 45°, b_k + 45°)$. A rotation by 45° effectively swaps the roles of the transmitted and reflected detectors. Each party, when measuring on the rotated basis, needs to apply a NOT operation to the measurement outcome. The results of these four measurements are combined into two binary files, $x(a_j, b_k)$ and $y(a_j, b_k)$, by interleaving their respective bits. In order to obtain long enough strings for a stable compression, see figure 3, this measurement is repeated 11 times and the results concatenated, obtaining strings of average length 10^5 bits.

For each angle of separation θ, we measure four basis (a_0, b_0), (a_1, b_0), (a_1, b_1), and (a_0, b_1), then calculate the NCD between $x(a_j, b_k)$ and $y(a_j, b_k)$ using Eq. (8), in order to obtain the value of S.

5. Results

The inequality is experimentally tested by evaluating S in Eq. (9) for a range of θ; the results [points (c), (d) in figure 2] are consistently lower than the trace (a) calculated via entropy using Eq. (7), and than a simulation with the same compressor (b). This is because the LZMA Utility is not working exactly at the Shannon limit, and also due to imperfect state generation and detection.

Although we set out in this work to avoid a statistical argument in the interpretation of measurement results, we do resort to statistical techniques to assess the confidence in an experimental finding of a violation of inequality (9). To estimate an uncertainty of the experimentally obtained values for S, we set $\theta = 8.6°$, for which we expect the maximum violation, and collected results from a larger number of photon pairs. We then repeated the measurement of S, as described in the previous section, 10 times, and considered the average value and standard deviation of this set obtaining the final result of $S(\theta = 8.6°) = 0.0494 \pm 0.0076$.

6. Discussion

There is a trend to look at physical systems and processes as programs run on a computer made of the constituents of our universe. Although this point of view has been already extensively used in quantum information theory, we present a complementary algorithmic approach for an explicit, experimentally testable example. This algorithmic approach is complementary to the orthodox Bell inequality approach to quantum nonlocality [1] that is statistical in its nature.

The Kolmogorov complexity of the output of UTM must obey distance properties as shown in [7, 4], and can be approximated by compression. The distance properties lead to inequality (9), which we find violated in the specific case of polarization-entangled photon pairs. Therefore, no hidden variables can be encoded as programs for spatially separated UTMs.

We would like to stress that our analysis of the experimental data is purely and consistently algorithmic. We do not resort to statistical methods that are alien to the concept of computation. In addition, the algorithmic approach does not use the notion of an ensemble and the assumption that each bit in a data string comes from an i.i.d. source. The compression treats the string of data as a single entity, and does not ignore correlations between subsequent string elements.

We have become aware of a recent arXive submission by Wolf [12] where this algorithmic approach is used to provide a different viewpoint on nonlocality that does not require counterfactual reasoning.

Acknowledgments

We acknowledge the support of this work by the National Research Foundation & Ministry of Education in Singapore, partly through the Academic Research Fund MOE2012-T3-1-009. P.K. and D.K. are also supported by the Foundational Questions Institute (FQXi). A.C. also thanks Andrea Baronchelli for the suggestions on the use of compression software.

[1] Bell J S 1964 *Physics* **1** 195–200
[2] Braunstein S L and Caves C M 1988 *Phys. Rev. Lett.* **61** 662–665 URL http://link.aps.org/doi/10.1103/PhysRevLett.61.662
[3] Shannon C E 1948 *Bell System Technical Journal* **27** 379–423 URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024
[4] Cilibrasi R and Vitányi P M B 2005 *Information Theory, IEEE Transactions on* **51** 1523–1545

[5] Fuchs C A 1992 Algorithmic information theory and the hidden variable question *Workshop on Squeezed States and Uncertainty Relations* NASA Conference Publication 3135 ed Han D, Kim Y S and Zachary W W (NASA, Washington, DC) pp 83–85
[6] Gill R D 2003 Time, finite statistics, and Bell's fifth position *Proc. of "Foundations of Probability and Physics - 2"* (*Ser. Math. Modelling in Phys., Engin., and Cogn. Sc.* vol 5 2002) (Växjö Univ. Press) pp 179–206
[7] Li M, Chen X, Li X, Ma B and Vitányi P M B 2004 *Information Theory, IEEE Transactions on* **50** 3250–3264
[8] Santos E 1986 *Physics Letters A* **115** 363–365
[9] Schumacher B W 1991 *Phys. Rev. A* **44** 7047–7052
[10] Kurzyński P and Kaszlikowski D 2014 *Phys. Rev. A* **89** 012103
[11] Wajs M, Kurzyński P and Kaszlikowski D 2015 *Phys. Rev. A* **91** 012114
[12] Wolf S 2015 *arXiv* (*Preprint* `1505.07037`) URL `http://arxiv.org/abs/1505.07037`
[13] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd ed (Wiley-Interscience) ISBN 9780471748816
[14] Pavlov I http://www.7-zip.org/sdk.html URL `http://www.7-zip.org/sdk.html`
[15] Seward J http://www.bzip.org/ URL `http://www.bzip.org`
[16] Gailly J l and Adler M http://www.gzip.org/ URL `http://www.gzip.org/`
[17] Welch T 1984 *Computer* **17** 8–19 ISSN 0018-9162
[18] MATLAB R2010a, The MathWorks, Inc., Natick, Massachusetts, United States.
[19] Matsumoto M and Nishimura T 1998 *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **8** 3–30
[20] Marsaglia G and Tsang W W 2000 *Journal of Statistical Software* **5** 1–7 ISSN 1548-7660 URL `http://www.jstatsoft.org/v05/i08`
[21] Kwiat P G, Mattle K, Weinfurter H, Zeilinger A, Sergienko A V and Shih Y 1995 *Phys. Rev. Lett.* **75** 4337–4341 URL `http://link.aps.org/doi/10.1103/PhysRevLett.75.4337`
[22] Kurtsiefer C, Oberparleiter M and Weinfurter H 2001 *Phys. Rev. A* **64** 023802
[23] Clauser J F, Horne M, Shimony A and Holt R 1969 *Phys. Rev. Lett.* **23** 880–884