Probing the quantum-classical boundary with compression software

Hou Shun Poh[1], **Marcin Markiewicz**~~[2,3]~~[2], **Paweł Kurzyński**~~[1,4]~~[1,3], **Alessandro Cerè**[1], **Dagomir Kaszlikowski**~~[1,5]~~[1,4] **and Christian Kurtsiefer**~~[1,5]~~[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 ~~experimentally demonstrate an impossibility to reproduce quantum bipartite correlations with a deterministic universal Turing machine. We use the Normalized Information Distance (NID) that allows the comparison of two pieces of data without detailed knowledge about their origin. Using NID,~~ 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 ~~for output of two local deterministic universal Turing machines with correlated inputs. This inequality is violated by correlations generated by a maximally entangled polarization two-photon state. The violation is shown using a freely available lossless compressor. The presented technique may allow to complement the common statistical interpretation of quantum physics by an algorithmic one that does not require the assumption of an~~ ~~*independent identically distributed* (~~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 polarisation-entengled photonic pairs and readily available compression software. We argue that our approach relaxes the i.i.d. ~~) realization of photon pairs~~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

~~The idea that physical processes can be considered as computations done on some universal machines traces back to Turing and von Neumann [1]. This resulted in a completely new approach to science in which the complexity of observed phenomena is closely related to the complexity of computational resources needed to simulate them [2]. There are physical phenomena that cannot be traced with analytical tools, which further motivated a computational approach to physics [3]. Moreover, the idea of quantum computation [4] led to a discovery of a few problems efficiently traceable on quantum computers but not on classical ones [5, 6]~~

In a standard Bell scenario [18] 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.

In order to calculate correlations one has to estimate probabilities $p(x, y|a, b)$ that Alice outcome is x and Bob's is y, given the measurement settings are a and b, respectively. This is done under the assumption that Alice and Bob perform their measurements on independent and identically distributed (i.i.d.) pairs of non-classically correlated systems and therefore one has $p(x, y|a, b) = N(x, y|a, b)/N(a, b)$. $N(x, y|a, b)$ is the number of times outcomes x and y detected 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 80's by Braunstein and Caves [19]. Instead of using correlation functions they constructed a test of local realism using conditional Shannon entropies $H(a|b) = H(ab) - H(b)$, where $H(x) = -\sum_i p(x = x_i)\log_2 p(x = x_i)$. Although these inequalities are not tight their immediate advantage is that one does not need to worry about the labelling and they work for any number of outcomes. Instead, one is interested in a more fundamental problem, namely how much information about Alice's outcomes is contained in Bob's ones.

The ~~question arises if the complexity of the output of~~ simplest information-theoretic Bell inequality is of the following form

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

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

Although the method of Braunstein and Caves offers a conceptually new approach, it still requires estimation of probabilities $p(x, y|a, b)$. Therefore, actual experimental implementations of such information-theoretic tests are akin to the standard ones because they require an identical statistical analysis of data strings obtained by Alice and Bob. However, Shannon entropy of a ~~system can be used as a signature of its non-classicality. In this paper we show that there are processes which cannot be reproduced on local universal Turing machines (UTM) at all, independently of the available classical resources, following a similar approach by Fuchs [7]. We first revisit~~

~~the concept of Kolmogorov complexity, a measure of the classical complexity of a phenomenon, and later apply it to derive a bound on classical descriptions [9].Next, we use the fact that Kolmogorov complexity can be approximated by compression algorithms [10]. We then compress experimental data obtained from polarization measurements on entangled photon pairs and show a violation of the classical bound.~~ 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 [20]. 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.

1.1. ~~*Kolmogorov complexity*~~

~~Consider the description of a machine, whether classical or quantum, that outputs a string x of 0's and 1's. In the case of a UTM, we can always write a program Λ that generates x. The simplest such program is obviously 'PRINT x'~~For most data strings it is hard, or impossible, to prove that a compression algorithm we found is the best. However, this ~~is not optimal: in many cases the program can be much shorter than the string itself.~~

~~This brings us to the~~ 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 ~~$K(x)$,~~ 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 ~~minimal length $l(\Lambda)$ of all programs~~ length $l(\Lambda)$ of the shortest program Λ~~that reproduce a specific output~~ , which, when fed into a UTM, produces an output string x~~. If $K(x)$ is comparable to the length of the output $l(x)$ then our algorithmic description of x is inefficient, and x is called algorithmically random [11]. In most cases~~ . $K(x)$ is ~~uncomputable [16].To circumvent this issue, we can estimate $K(x)$ with some efficient lossless compression~~ in general uncomputable, however realistic compression algorithms $C(x)$ ~~[10].~~bound it from above [10], i.e., $K(x) \le C(x)$. In addition, compression algorithms can be applied to data strings that are generated by non i.i.d. sources.

1.1. ~~*Bipartite systems*~~

~~We now extend this picture to~~ 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 entangled photonic pairs. 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 [7] and that an alternative approach to non i.i.d. sources was discussed by Gill [8].

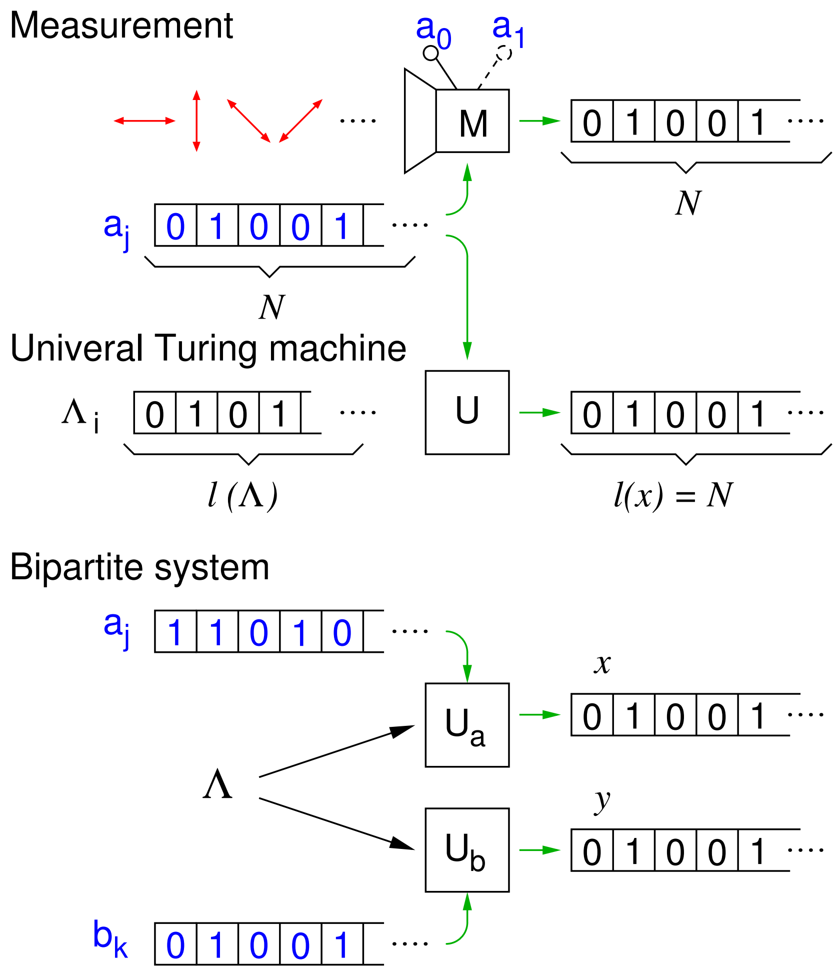

Figure 1. The figure and the caption has to be changed – see references to this figure in the text.

2. Algorithmic approach to Bell scenario

Let us start with the following assumption: *If there exists a local-realistic description of Bell measurements then it can be encoded as a program for UTMs that output bit strings corresponding to measurement outcomes of a realistic experiment.* In fact, this assumption says that the physical Church-Turing thesis applies to hypothetical local hidden variables. Such program would contain full information about a physical system that is necessary to compute the outcomes of measurements performed on it, provided a description of a measurement setting is also available to the UTM (see Fig. 1 top).

In particular, we consider two spatially separated ~~UTM's U_A (Alice) and U_B (Bob). If these machines cannot communicate, they generate two output strings that~~ UTMs: UTM$_A$ and UTM$_B$. Each UTM is fed with 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, one after another, on a single bit string which is an input into the machine.

The input programs to both machines are ~~independent, although the programs fed into the machines can be correlated. Moreover, the input programs are~~ classical bit strings ~~so~~ and the correlations between them ~~must be classical.~~

~~We determine the complexity of the strings using the *Normalized Information Distance* (NID) [9]. This distance compares two data sets without detailed knowledge about their origin. In practice, we evaluate an approximation to the NID, the *Normalized Compression Distance* (NCD) [10], using a lossless compression software, in our case the LZMA Utilities, based on the Lempel-Ziv-Markov chain algorithm [17].~~

~~We consider a model experiment, similar to the Bell test [18]: a source emits pairs of photons traveling to two separate polarization analyzers M~~result only from the shared information about the preparation (Λ). In addition, UTM$_A$ ~~(Alice) and M~~has no access to a description of a measurement setting b_k that was fed to UTM$_B$ ~~(Bob). Each analyzer has two outputs labeled 0~~ and vice versa. Therefore, the output strings x_j and ~~1, and can be set along directions a_0 or a_1 for M$_A$, and b_0 or b_1 for M$_B$. The analyzers' outputs are bit strings (see 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 or 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.~~ 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 posssible hidden variable must be encoded in Λ.

The ~~output x of each analyzer can be described as the output of a UTM, fed with the settings~~ simulation is done in the following way (see Fig. 1 bottom). One copy of a program Λ is sent to Alice and another one to Bob. The first copy is concatenated with a description of a measurement setting a_j ~~or~~ and fed into UTM$_A$, whereas the second one is concatenated with b_k ~~, and a program Λ, which contains the information about generating the correct output for every detection event and for every setting. For a string of finite length $l(x) = N$, Λ has to describe the 4^N possible events. The length of the shortest Λ is the Kolmogorov complexity of the generated string. Next, we describe the output of the experiment as the output of two local non-communicating UTMsU_A~~ and ~~U_B. We feed Λ to both of them and obtain two output strings~~ fed into UTM$_B$. At this point both UTMs start to compute and after some time the output strings x_j and y_k are produced. Next, we repeat the whole procedure but with different measurement settings. The goal is to produce all four pairs of bit strings: $\{x_0, y_0\}$, ~~x~~ $\{x_0, y_1\}$, $\{x_1, y_0\}$ and ~~y, both of length N. The program has to describe the behavior of all $2N$ events for all possible settings a_j and b_k, hence 16^N possibilities.~~ $\{x_1, y_1\}$.

~~The Kolmogorov complexity $K(x, y)$ of two bit stringsis the~~

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 on two spatially separated UTMs. The general idea is to compare the two bit strings using some mathematical tools. In the standard approach [18] 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 $x_k^{(n)}$. 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 ~~them simultaneously. $K(x,y)$ can be shorter than $K(x) + K(y)$~~ 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. Note, that if x and y are correlated ~~- the more correlated they are , the simpler it is to compute one string knowing the other.~~

~~A distance measure between~~ one may have $K(x,y) < K(x) + K(y)$, 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 ~~y called~~ $K(y_k)$. 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) ~~was introduced in[9]:~~ introduced in [9]

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

~~The NID~~ This distance is a metric ~~and thus obeys the triangle inequality~~

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

~~It holds~~ up to a correction ~~of order $\log(l(x))$, which can be neglected for sufficiently long strings [9].~~

2.2. ~~*Information Inequality*~~

~~We consider bit strings x_{a_j} and y_{b_k} generated by Alice and Bob with fixed settings a_j~~ ~~$\frac{\log_2 N}{N}$~~, where N is the length of strings x and ~~b_k. Inequality (3) transforms to~~ 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. Note, that the metric approach to Bell inequalities was used before in [12, 13, 14, 15]. The NID obeys the triangle inequality

$$\mathrm{NID}(\mathrm{x}_{\mathrm{a}_0}, \mathrm{y}_{\mathrm{b}_0}) + \mathrm{NID}(\mathrm{y}_{\mathrm{b}_0}, \mathrm{y}_{\mathrm{b}_1}\mathrm{z}) \geq \mathrm{NID}(\mathrm{x}_{\mathrm{a}_0}, \mathrm{y}_{\mathrm{b}_1}\mathrm{z})\,. \tag{3}$$

~~However, NID($\mathrm{y}_{\mathrm{b}_0}, \mathrm{y}_{\mathrm{b}_1}$)~~ In our case we can write the following

$$\mathrm{NID}(\mathrm{x}_0, \mathrm{y}_0) + \mathrm{NID}(\mathrm{y}_0, \mathrm{y}_1) \geq \mathrm{NID}(\mathrm{x}_0, \mathrm{y}_1)\,. \tag{4}$$

However, NID($\mathrm{y}_0, \mathrm{y}_1$) cannot be determined experimentally because the strings ~~y_{b_0} and y_{b_1} come from measurements of incompatible observables.~~ 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 derivation of all known Bell inequalities (with a possible exception in [30]). 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.

We therefore use the triangle inequality ~~$\mathrm{NID}(\mathrm{x}_{\mathrm{a}_1}, \mathrm{y}_{\mathrm{b}_0}) + \mathrm{NID}(\mathrm{x}_{\mathrm{a}_1}, \mathrm{y}_{\mathrm{b}_1}) \geq \mathrm{NID}(\mathrm{y}_{\mathrm{b}_0}, \mathrm{y}_{\mathrm{b}_1})$,~~ $\mathrm{NID}(\mathrm{x}_1, \mathrm{y}_0) + \mathrm{NID}(\mathrm{x}_1, \mathrm{y}_1) \geq \mathrm{NID}(\mathrm{y}_0, \mathrm{y}_1)$, and combine it with (4) to get:

$$\mathrm{NID}(\mathrm{x}_{\mathrm{a}_0 0}, \mathrm{y}_{\mathrm{b}_0 0}) + \mathrm{NID}(\mathrm{x}_{\mathrm{a}_1 1}, \mathrm{y}_{\mathrm{b}_0 0}) + \mathrm{NID}(\mathrm{x}_{\mathrm{a}_1 1}, \mathrm{y}_{\mathrm{b}_1 1}) \geq \mathrm{NID}(\mathrm{x}_{\mathrm{a}_0 0}, \mathrm{y}_{\mathrm{b}_1 1})\,. \tag{5}$$

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

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

$$S' = \mathrm{NID}(\mathrm{x}_{\mathrm{a}_0 0}, \mathrm{y}_{\mathrm{b}_1 1}) - \mathrm{NID}(\mathrm{x}_{\mathrm{a}_0 0}, \mathrm{y}_{\mathrm{b}_0 0}) - \mathrm{NID}(\mathrm{x}_{\mathrm{a}_1 1}, \mathrm{y}_{\mathrm{b}_0 0}) - \mathrm{NID}(\mathrm{x}_{\mathrm{a}_1 1}, \mathrm{y}_{\mathrm{b}_1 1}) \leq 0 \tag{6}$$

The violation of local realism occurs if S' is positive.

To test ~~this inequality,~~ the positivity of S' we have to address ~~the following problem. We~~ a problem that also appears in standard Bell scenarios (for a suggested resolution see [30]). For example, we can set up a source to generate entangled photon pairs in an arbitrary state~~but~~. However, for every experimental run i, with the same preparation~~the resulting string x_{i,a_j}~~, the generated local string x_{j_i} can be different. Consequently, the corresponding program Λ_i ~~is~~ can be different for every experimental run. Therefore, we need to argue why $NID(x_{j_i}, y_{0_i})$ and $NID(x_{j_{i'}}, y_{1_{i'}})$ can be considered simply as $NID(x_j, y_0)$ and $NID(x_j, y_1)$.

~~It is reasonable to assume~~ To address this issue, we introduce a new assumption that 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 experimetnally. ~~Without these assumptions the same physical preparation of the experiment has different consequences and thus the notion of preparation loses its meaning. More generally, the predictive power of science can be stated as: the same preparation results in the same complexity of observed phenomena.~~

3. ~~Estimation of Kolmogorov complexity~~

~~In~~ Another problem that needs to be resolved is that in general the 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 (1). We will briefly discuss it, however since we aim at algorithmic approach we later focus on the other one.

2.1. ~~Statistical Approach~~

2.0.1. Statistical approach The statistical approach uses an ensemble X of all possible 2^N bit strings of length N ~~bit strings~~ and looks for their average Kolmogorov complexity. The ensemble average is the Shannon entropy $H(X)$ [16] and thus

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

Inequality (5) becomes an entropic Bell inequality by Braunstein and Caves [19] if local entropies are maximal, i.e., $H(x) = H(y) = N$. They 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 $\theta = 8.6°$. ~~Plots of S versus angle of separation θ. (a) Result obtained from (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 $\theta = 8.6°$. This statistical approach requires the assumption that output of the systems are *independent identically distributed* (i.i.d.).~~

2.1. ~~Algorithmic approach~~

2.0.1. Algorithmic approach On the other hand, it is possible to avoid a statistical description of our experiment following Ref. [10] where it was shown that the Kolmogorov complexity can be well approximated by compression algorithms. In [10] the *Normalized Compression Distance* (NCD) is introduced

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

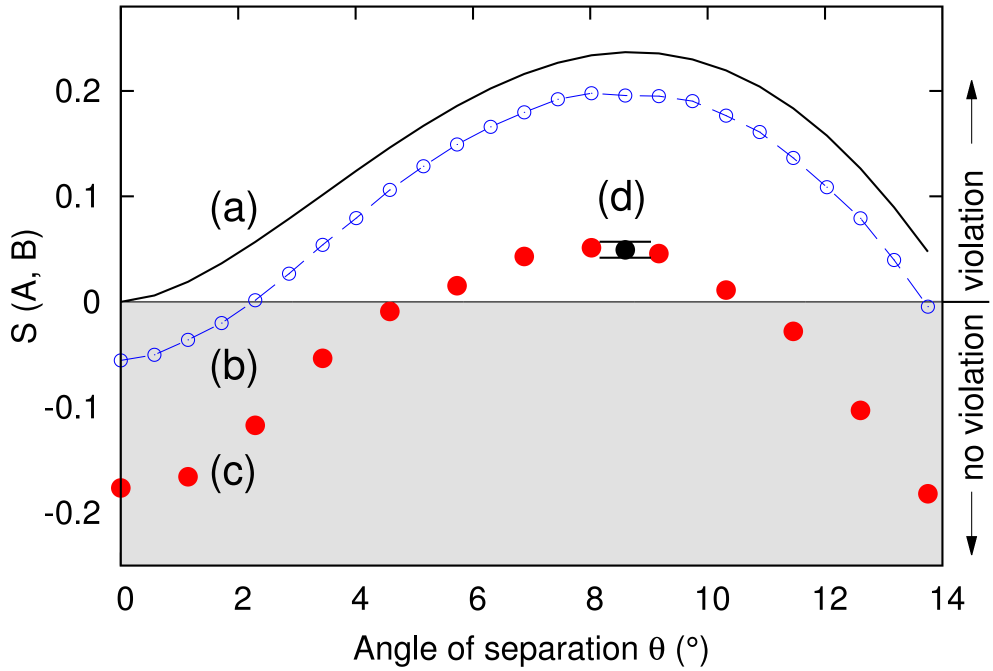

Figure 2. Plots of S versus angle of separation θ. (a) Result obtained from (7), (b) result obtained from using the LZMA compressor on numerically generated data, (c) measurement of S in the experiment shown in figure 5, and (d) longer measurement at the optimal angle $\theta = 8.6°$.

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 (6) leads to a new inequality:

$$S' \to S = \mathrm{NCD}(\mathrm{x}_{\mathrm{a}_0}, \mathrm{y}_{\mathrm{b}_1}) - \mathrm{NCD}(\mathrm{x}_{\mathrm{a}_0}, \mathrm{y}_{\mathrm{b}_0}) - \mathrm{NCD}(\mathrm{x}_{\mathrm{a}_1}, \mathrm{y}_{\mathrm{b}_0}) - \mathrm{NCD}(\mathrm{x}_{\mathrm{a}_1}, \mathrm{y}_{\mathrm{b}_0}) \leq 0\,. \quad (9)$$

This expression can be tested experimentally because the NCD is operationally defined. Moreover, it was shown in [10] that NCD is a metric up to a term $\frac{\log_2 N}{N}$, therefore () holds up to the same term.

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 [20]. This is necessary to ensure that it does not introduce any unintended artifacts that lead to

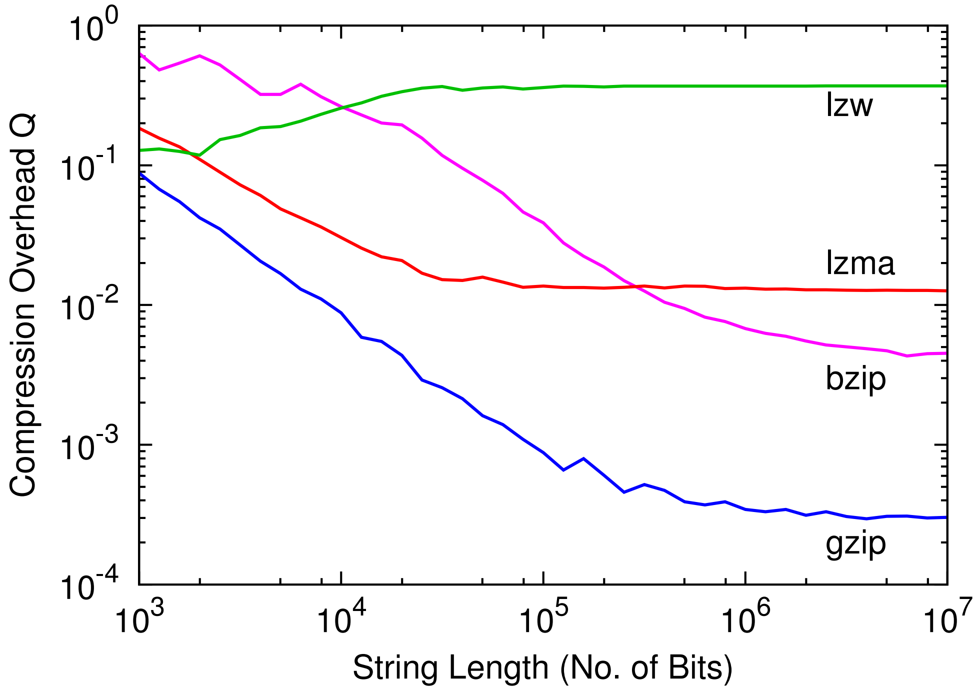

Figure 3. Comparison of the compression overhead Q obtained using four different compression algorithms on pseudo-random strings of varying lengths. The expected value for an ideal compressor is 0. From this characterization we can exclude LZW as a useful compressor for our application.

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 [17], BZIP2 [21], GZIP [22] and LZW [23]. To eliminate the overhead associated with the compression of ASCII text files, we save data in a binary format.

For this characterization and a 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* [24] 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 [25], with ziggurat [26] as the algorithm that generates the required probability distribution. The complexity of this (deterministic) source of pseudorandom numbers should be high enough to *n*ot 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 Fig. 3, it can be seen that

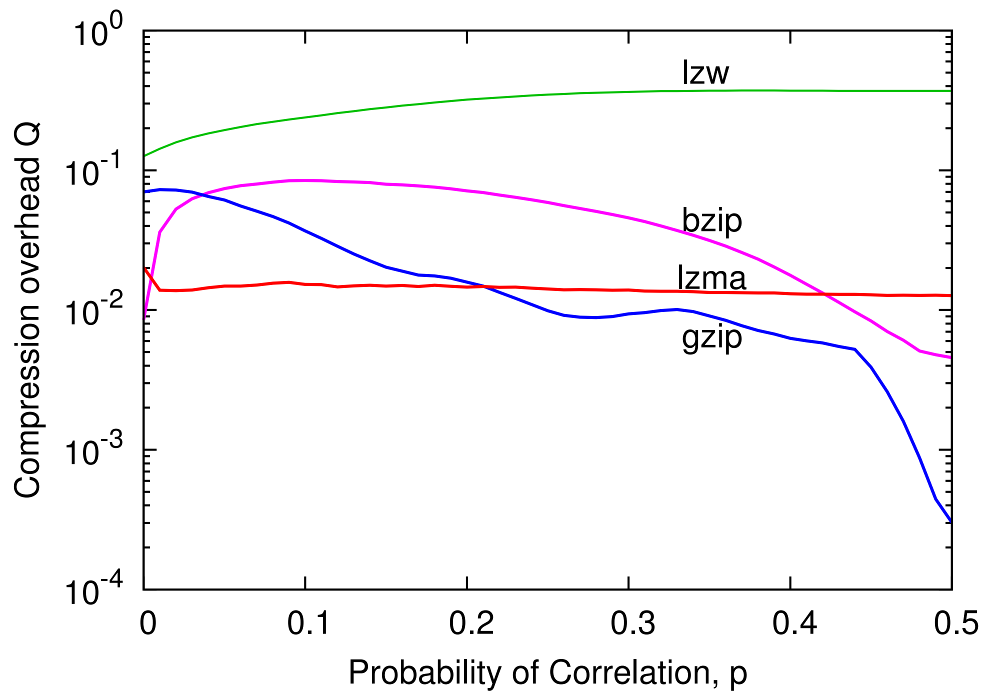

Figure 4. Compression overhead Q for the string xy as a function of the probability of pairwise correlation p between the bits of the generating strings x and y for three different compressors: BZIP2, GZIP, and LZMA.

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. The LZW compressor fails this test, converging to a Q of 0.37 for long string, while BZIP2, GZIP and LZMA give a Q below 10^{-1}.

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.

The two strings x and y are then 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. We then compress xy and evaluate the compression overhead Q as a function of p. The results for different compressors are shown in Fig. 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. 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 [10]. 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

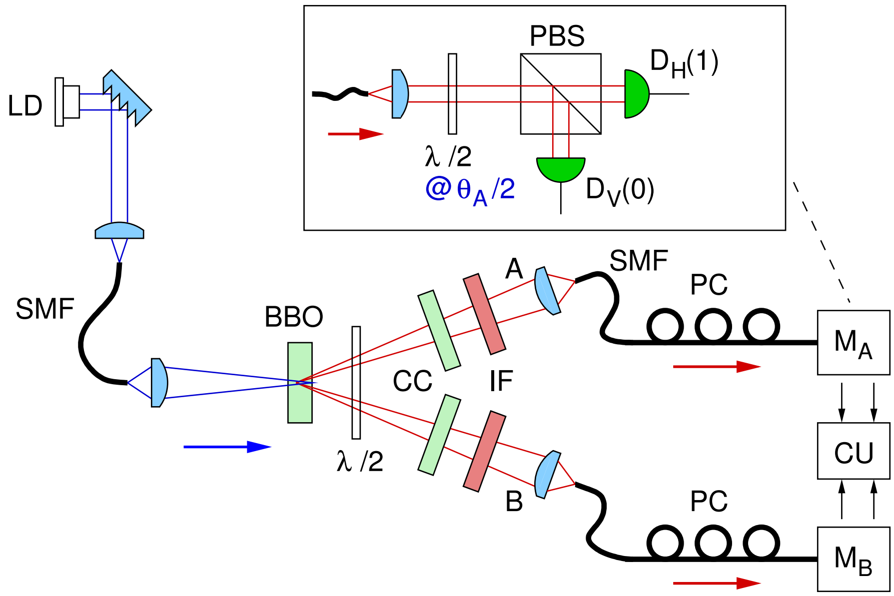

Figure 5. Schematic of the experimental set-up. Polarization correlations of entangled-photon pairs are measured by the polarization analyzers M_A and M_B, each consisting of a half wave plate ($\lambda/2$) followed by a polarization beam splitter (PBS). All photons are detected by Avalanche photodetectors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

tests.

The numerical simulation also verifies the angle that maximizes the violation of (2.0.1). 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 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 [27]. 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 [28]. 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 (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 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 [29] 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

We record coincidences of detection events between detectors at A and B. For each PBS, the transmitted output is associated with 0 and the reflected one with 1. The resulting binary strings x from A, and y from B are written into two individual binary files. From these, we calculate the NCD using (8). This procedure is repeated for each of the four settings (a_0, b_0), (a_1, b_0), (a_1, b_1), and (a_0, b_1) in order to obtain the value for S.

4.2. Symmetrization of detector efficiencies

To remove the bias due to differences in the detection efficiency of the APDs in the experiment, we also measure for each setting the associated orthogonal ones. The experimental setup (see figure 5) uses four APDs: D_{HA}, D_{VA} (Alice), and D_{HB}, D_{VB} (Bob) to register photon pair events in the two spatial modes. By denoting events at D_H and D_V as 1 and 0, the four possible output patterns are 00, 01, 10, and 11, where the least and most significant bit corresponds to the Alice and Bob mode, respectively. Due to differences in the the losses in the transmitted and reflected port of the PBS, efficiencies in coupling light into the APDs, and the quantum efficiencies of APDs, the detection efficiencies for the four output combinations are different. The resulting effective pair efficiencies are then given by the product of the contributing detection efficiencies η_{VB}, η_{HB}, η_{VA}, and η_{HA}.

This asymmetry will skew the statistics of the measurement results. We symmetrize the effective pair efficiencies for each (θ_A, θ_B) measuring also the following settings for the half wave plates: $(\theta_A + 45°, \theta_B)$, $(\theta_A, \theta_B + 45°)$, and $(\theta_A + 45°, \theta_B + 45°)$. This procedure swaps the output ports of the PBS at which each polarization state is detected. The resulting outcomes are then interleaved, providing an uniform detection probability for the four possible outcomes. The effective pair detection efficiency for all four combinations is then $(\eta_{VB}\eta_{VA} + \eta_{VB}\eta_{HA} + \eta_{HB}\eta_{VA} + \eta_{HB}\eta_{HA})/4$.

5. Results

The inequality is experimentally tested by evaluating S in (2.0.1) for a range of θ; the results [points (c), (d) in figure 2] are consistently lower than the trace (a) calculated via entropy using (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.

For $\theta = 8.6°$ we collected results from a large number of photon pairs. 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 (2.0.1). To estimate an uncertainty of the experimentally obtained values for S, this large data set was subdivided into files with length greater than 10^5 bits. The results from all these files are then averaged to obtain the final result of $S(\theta = 8.6°) = 0.0494 \pm 0.0076$, with the latter indicating a relatively small standard deviation over these different subsets.

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. ~~We show that this is not possible if one uses a computation paradigm of a local UTM. Although this~~ Although this point of view has been already extensively ~~researched~~ 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 [18] that is statistical in its nature.

The Kolmogorov complexity of the output of ~~local~~ UTM must obey distance properties as shown in [9, 10], and can be approximated by compression. The distance properties lead to inequality (2.0.1), which we find violated in the specific case of polarization-entangled photon pairs. Therefore, ~~at least this physical processes can not~~ no hidden variables can be encoded as programs ~~on local~~ for spatially seprated 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. ~~assumption~~source. The compression treats the string of data as a single entity, and does not ignore correlations between subsequent string elements. ~~Our approach allows us therefore to omit the notion of probability, at least for the case at hand. If it can be extended to other quantum experiments, it would offer an alternative with less assumptions to the commonly used statistical interpretation of quantum theory.~~

We have become aware of a recent ~~work~~ arXive submission by Wolf [30] ~~inspired by the ideas presented in this work,~~ 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 hints on the use of compression software.

[1] Coveney P V and Highfield R 1996 *Frontiers of Complexity: The Search for Order in a Chaotic World* (Fawcett Columbine) ISBN 9780449910818

[2] Wolfram S 1985 *Phys. Rev. Lett.* **54** 735–738 URL `http://link.aps.org/doi/10.1103/PhysRevLett.54.735`

[3] Moore C 1990 *Phys. Rev. Lett.* **64** 2354–2357 URL `http://link.aps.org/doi/10.1103/PhysRevLett.64.2354`

[4] Feynman R P 1982 *Int. J. Theor. Phys.* **21** 467–488 URL `http://link.springer.com/10.1007/BF02650179`

[5] Shor P W 1994 Algorithms for quantum computation: discrete logarithms and factoring *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (Santa Fe, NM: IEEE) pp 124–134 ISBN 0818665807

[6] Grover L K 1996 A fast quantum mechanical algorithm for database search *Proceedings, 28th Annual ACM Symposium on the Theory of Computing* p 212 URL `http://arxiv.org/abs/quant-ph/9605043`

[7] 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

[8] 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

[9] Li M, Chen X, Li X, Ma B and Vitányi P M B 2004 *Information Theory, IEEE Transactions on* **50** 3250–3264

[10] Cilibrasi R and Vitányi P M B 2005 *Information Theory, IEEE Transactions on* **51** 1523–1545

[11] Li M and Vitanyi P M B 2009 *An Introduction to Kolmogorov Complexity and Its Applications* Texts in Computer Science (Springer New York) ISBN 9780387498201

[12] Santos E 1986 *Phys. Lett. A* **115** 363

[13] Schumacher B W 1991 *Phys. Rev. A* **44** 7047

[14] Kurzynski P Kaszlikowski D 2014 *Phys. Rev. A* **89** 012103

[15] Wajs M, Kurzynski P and Kaszlikowski D 2015 *Phys. Rev. A* **91** 012114

[16] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd ed (Wiley-Interscience) ISBN 9780471748816

[17] Pavlov I http://www.7-zip.org/sdk.html URL `http://www.7-zip.org/sdk.html`

[18] Bell J S 1964 *Physics* **1** 195–200

[19] 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`

[20] Shannon C E 1948 *Bell System Technical Journal* **27** 379–423 URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024`

[21] Seward J http://www.bzip.org/ URL `http://www.bzip.org`

[22] Gailly J l and Adler M http://www.gzip.org/ URL `http://www.gzip.org/`

[23] Welch T 1984 *Computer* **17** 8–19 ISSN 0018-9162

[24] MATLAB R2010a, The MathWorks, Inc., Natick, Massachusetts, United States.

[25] Matsumoto M and Nishimura T 1998 *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **8** 3–30

[26] Marsaglia G and Tsang W W 2000 *Journal of Statistical Software* **5** 1–7 ISSN 1548-7660 URL `http://www.jstatsoft.org/v05/i08`

[27] 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`

[28] Kurtsiefer C, Oberparleiter M and Weinfurter H 2001 *Phys. Rev. A* **64** 023802

[29] Clauser J F, Horne M, Shimony A and Holt R 1969 *Phys. Rev. Lett.* **23** 880–884

[30] Wolf S 2015 *arXiv* (*Preprint* `1505.07037`) URL `http://arxiv.org/abs/1505.07037`