Device-independent information protocols:
Measuring dimensionality, randomness and nonlocality

Ph.D. Thesis

Ph.D. Candidate:
Rodrigo Gallego

Thesis Supervisor:
Dr. Antonio Acín

ICFO - Institut de Ciències Fotòniques

Contents

Abstract

The device-independent formalism is a set of tools to analyze experimental data and infer properties about systems, while avoiding almost any assumption about the functioning of devices. It has found applications both in fundamental and applied physics: some examples are the characterization of quantum nonlocality and information protocols for secure cryptography or randomness generation. This thesis contains novel results on these topics and also new applications such as device-independent test for dimensionality.

After an introduction to the field, the thesis is divided in four parts. In the first we study device-independent tests for classical and quantum dimensionality. We investigate a scenario with a source and a measurement device. The goal is to infer, solely from the measurement statistics, the dimensionality required to describe the system. To this end, we exploit the concept of dimension witnesses. These are functions of the measurement statistics whose value allows one to bound the dimension. We study also the robustness of our tests in more realistic experimental situations, in which devices are affected by noise and losses. Lastly, we report on an experimental implementation of dimension witnesses. We conducted the experiment on photons manipulated in polarization and orbital angular momentum. This allowed us to generate ensembles of classical and quantum systems of dimension up to four. We then certified their dimension as well as its quantum nature by using dimension witnesses.

The second part focuses on nonlocality. The local content is a nonlocality quantifier that represents the fraction of events that admit a local description. We focus on systems that exhibit, in that sense, maximal nonlocality. By exploiting the link between Kochen-Specker theorems and nonlocality, we derive a systematic recipe to construct maximally nonlocal correlations. We report on the experimental implementation of correlations with a high degree on nonlocality in comparison with all previous experiments on nonlocality. We also study maximally nonlocal correlations in the multipartite setting, and show that the so-called GHZ-state can be used to obtain correlations

suitable for multipartite information protocols, such as secret-sharing.

The third part studies nonlocality from an operational perspective. We study the set of operations that do not create nonlocality and characterize nonlocality as a resource theory. Our framework is consistent with the canonical definitions of nonlocality in the bipartite setting. However, we find that the well-established definition of multipartite nonlocality is inconsistent with the operational framework. We derive and analyze alternative definitions of multipartite nonlocality to recover consistency. Furthermore, the novel definitions of multipartite nonlocality allows us to analyze the validity of information principles to bound quantum correlations. We show that 'information causality' and 'non-trivial communication complexity' are insufficient to characterize the set of quantum correlations.

In the fourth part we present the first quantum protocol attaining full randomness amplification. The protocol uses as input a source of imperfect random bits and produces full random bits by exploiting nonlocality. Randomness amplification is impossible in the classical regime and it was known to be possible with quantum system only if the initial source was almost fully random. Here, we prove that full randomness can indeed be certified using quantum non-locality under the minimal possible assumptions: the existence of a source of arbitrarily weak (but non-zero) randomness and the impossibility of instantaneous signaling. This implies that one is left with a strict dichotomic choice regarding randomness: either our world is fully deterministic or there exist events in nature that are fully random.

List of publications

This thesis is mainly based on the following publications.

- R. Gallego, N. Brunner, C. Hadley and A. Acín. Device-Independent Tests of Classical and Quantum Dimensions. *Phys. Rev. Lett.* **105,** 230501 (2010).

- R. Gallego, L. Würflinger, A. Acín. and M. Navascués. Quantum Correlations Require Multipartite Information Principles *Phys. Rev. Lett.* **107,** 210403 (2011).

- R. Gallego, L. Würflinger, A. Acín. and M. Navascués. Operational Framework for Nonlocality *Phys. Rev. Lett.* **109,** 070401 (2012).

- L. Aolita, R. Gallego, A. Acín, A. Chiuri, G. Vallone, P. Mataloni and A. Cabello. Fully nonlocal quantum correlations *Phys. Rev. A* **85,** 032107 (2012).

- L. Aolita, R. Gallego, A. Cabello and A. Acín. Fully Nonlocal, Monogamous, and Random Genuinely Multipartite Quantum Correlations *Phys. Rev. Lett.* **108,** 100401 (2012).

- M. Hendrych, R. Gallego, M. Micuda, N. Brunner, A. Acín, and J. P. Torres. Experimental estimation of the dimension of classical and quantum systems *Nat. Phys.* **8,** 588–591 (2012).

- M. Dall'Arno, E. Passaro, R. Gallego and A. Acín. Robustness of device-independent dimension witnesses *Phys. Rev. A* **86,** 042312 (2012).

- M. Dall'Arno, E. Passaro, R. Gallego, M. Pawlowski and A. Acín. Attacks on semi-device independent quantum protocols arXiv:1210.1272 (2012).

- R. Gallego, L. Masanes, G. De la Torre, C. Dhara, L. Aolita, and A. Acín, Full randomness from arbitrarily deterministic events. arXiv:1210.6514 (2012).

Chapter 1

Introduction

A measurement process is an interaction among physical systems from which one acquires classical information. These processes are one of the building elements of the development of scientific knowledge, as the results of measurements are to be confronted with theoretical models predicting the behavior of nature. It is often the case that the measurement process itself can be divided into sub-blocks. One may have a well-tested model for some of the sub-blocks, however other sub-blocks are to be tested experimentally. Consider for example the interaction of an atom and a magnetic field created by an electrical current. One may have a well-tested model for the generation of the field by the current, however the interaction between atom and field is to some extent uncharacterized. In this situation it is perfectly reasonable to make use of the well-tested model for the current producing the field when the results of the whole experiment are interpreted. Indeed, it is very useful to extend the explicative power of well-tested models as far as possible, so one can isolate the sub-block that is to be understood.

It is also common that the model under test is formulated within a theoretical framework assumed to be correct. To follow with the previous example, one can measure the value of the magnetic dipole of the atom by using a well-tested framework as Maxwell equations. Indeed, it seems reasonable to say that the more assumptions on the theoretical framework, the more precise the interpretation of the results. To summarize: measurements are a complex network of interacting processes. The interpretation of measurement results relies on very intricate assumptions about the functioning of the devices, and these assumptions enhance a deeper understanding of the measurement outcomes.

However, there exist situations where it is convenient to avoid as many

assumptions as possible about the internal working of the measurement device, even if one has a well-tested theory for some of its parts. We identify three possible scenarios in which this is the case: (i) When the unknown physical property to be measured is very fundamental, and therefore it is difficult to build theoretical models that do not rely on this unknown property. For example, if one aims at measuring the dimensionality of a system, it is difficult to build models that are dimension-free, that is, that do not make any initial assumption on the dimension. (ii) When the experiment is designed to certify a property of the theoretical framework describing the experiment. For example, Bell's theorem shows that there is no local theory able to describe measurements on certain quantum states. (iii) When one aims at performing a task in which making assumption is explicitly undesired. For example if the task has to be carried out in competition with a malicious agent that may take advantage of your incorrect assumptions.

In these scenarios one is compelled to interpret the measurement results by using as few assumptions as possible. This paradigm is known as device-independent and this thesis offers examples of its usefulness, both from the theoretical and the applied viewpoint. More interestingly, the interplay between scenarios (i), (ii) and (iii) makes that results initially conceived to make a fundamental statement about nature later find an application for certain device-independent tasks. Conversely, the study of device-independent tasks often yield to statements with a fundamental importance to understand our current theories. This thesis offers many examples of this interplay.

1.1 Dimensionality

The first chapter is devoted to device-independent tests for dimensionality. This fits in the scenario (i) that was considered above. The aim is to infer from the experimental data which is the dimensionality (the number of degrees of freedom) needed to describe a physical situation. Every model makes explicitly an assumption about the degrees of freedom, which is the very quantity to measure. Therefore, there is a dramatic restriction on the assumptions that one can use to describe the experiment. Nevertheless, it is reasonable to make some general assumptions about the theory that describes the experiment, for example, that it has to be compatible with the laws of quantum mechanics (or classical mechanics). Interestingly, by using such general assumptions it is possible to obtain relevant bounds on the dimensionality of the system from the raw experimental data. Furthermore, we provide an experimental demonstration of our results. That is, we

estimate the dimension of a physical system.

As mentioned above, in such tests, the only assumption is the theory describing the experiment (classical mechanics or quantum mechanics). This allows one to compare the performance of the two different theories for a fixed dimension. We show how to build tests for distinguishing quantum and classical systems of a fixed dimension.

Main results on dimensionality

Dimension witness

We derive a recipe to test the dimensionality of an arbitrary system only from the statistics of measurements performed on it. This method can be adapted to every experiment involving a source and a measurement device and all the theoretical machinery is distilled into a single linear combination of the observed statistics. We provide specific examples for classical and quantum systems useful in quantum information theory and provide a formalism to distinguish classical and quantum systems of a fixed dimension. Moreover, the techniques are appealing from an experimental viewpoint.

Detection loophole in tests for dimensionality

We study the performance of our tests for dimensionality in the presence of imperfect detectors. We derive lower and upper bounds on the detection efficiency necessary to bound the dimensionality of quantum and classical systems. Furthermore, we show that an extra assumption on the functioning of the devices allows one to bound the dimension for an arbitrary non-zero value of the detection efficiency.

Experimental demonstration of tests for dimensionality

We perform an experimental tests on photons up to dimension four. We exploit the polarization and angular momentum of photons and certify the dimension and the quantumness of the photons. This is the first experimental demonstration of a test of dimensionality.

1.2 Nonlocality

The second chapter focuses on nonlocality. The modern foundations for the study of nonlocality were settled by Bell's Theorem (also known as Bell inequalities). This theorem states that, according to the predictions

of quantum theory, the results of measurements performed on two spatially separated systems are not compatible with any local theory. In a sense, Bell designed an experiment that measures a property of any theory that aims at describing it, namely, whether it is nonlocal. This is by definition a scenario where it is mandatory to avoid as many assumptions as possible on the functioning of the device. Otherwise, one would end up with weak statements such: 'no local theory can describe the experiment, provided that such and such assumptions are true about the functioning of the devices'. That is, Bell theorem can be understood as a foundational example of the a device-independent tests where the scenario (ii) applies. As anticipated before, the impossibility of describing the experiment by a local theory has implications for device-independent tasks. In particular, nonlocality is intimately related to secrecy and randomness. This thesis contain several example of these relations.

Main results on nonlocality

Maximally nonlocal correlations

We design experiments to show that quantum mechanics is maximally nonlocal. As entanglement, nonlocality is not only a dichotomic feature and here exist quantifiers of nonlocality. It is often the case, as in Bell's original inequality, that quantum mechanics provides nonlocal correlations, but not as nonlocal as it would be possible within a non-signaling theory (a theory that does not allow for instantaneous transmission of information). We give a systematic recipe to construct maximally nonlocal correlations, that is, as nonlocal as allowed by the no-signaling principle. Furthermore, we perform an experimental implementation of these correlations, providing the most nonlocal correlations ever reported. Also we focus on maximally nonlocal correlations in a multiparty scenario. We find correlations that are maximally nonlocal, monogamous (*i.e.* uncorrelated with any other party, for example an eavesdropper) and locally random. We show that these correlations are suitable for device-independent cryptographic tasks such as secret sharing.

An operational framework for nonlocality

The second section focuses on operational frameworks for nonlocality and information principles. Here we build an operational framework to describe nonlocality analogous to the operational framework of LOCC (local operations and classical communication) for the study of entanglement. We show

that the standard definition of multipartite nonlocality adopted by the community is not consistent with the framework, and propose alternatives to recover consistency. These alternative consistent definitions allows us to study nonlocal correlations and its behavior under certain operations. We use all this machinery to analyze the performance of information-theoretic principles for quantum correlations. We show that the most promising information-theoretic principles (information causality and non-triviality of communication complexity) are insufficient to bound quantum correlations.

Certifying randomness by nonlocal correlations.

The fourth section focuses on randomness. Nonlocality is directly related to randomness: a nonlocal theory cannot be deterministic if compatible with the no-signaling principle. Therefore, Bell theorem can be used to certify that no deterministic theory can describe the experiment, or in other words, that the outputs are random. It has been shown by other authors that Bell experiments are randomness amplifiers. Our main result is that this amplification can be made arbitrarily large. That is, provided an arbitrarily small amount of randomness, we can certify an arbitrarily large randomness on the measurement outputs.

Chapter 2

Background

2.1 The device-independent formalism

In the device-independent formalism measurement processes are represented by two classical variables: the input and the ouput. The input $x \in \{1, \ldots, m\}$ codifies all the tunable parameters of the measurement device. One can imagine an apparatus with many knobs that modify the magnetic fields, the temperature, the happiness, or any other property that one believes interesting for an experiment. The precise way in which these knobs modify the parameter is irrelevant. One just needs to encode the position of the knobs into the variable x. The output $a \in \{1, \ldots, k\}$ represents the outcome of the measurement. Again, the relation of the outcome with a certain physical property is irrelevant in the formalism. The output a is an encoding of a certain variable that one is able to read from a pointer, a screen or any other output generator, see Fig 2.1.

It is usually assumed within the device-independent formalism that it is possible to prepare independent and identically distributed copies of the experiment (*i.i.d.* assumption). The ensemble of copies provides a set of experimental data comprising the input and output at every copy of the experiment, see Fig. 2.1. Under the *i.i.d.* assumption, one can compute $P(a|x)$, the probability of obtaining outcome a when the measurement labeled by x has been performed. This is done just by assigning probabilities in a frequentist manner, $P(a|x) = N(a, x)/N(x)$, where $N(\cdot)$ counts the number of events within the ensemble of copies [1]. After collecting those

[1] In real experimental situations, the set of data is obtained by reusing the same experimental device, rather than manufacturing many identical copies. By this procedure, the *i.i.d.* assumption is clearly not satisfied if the devices have memory [BCH+02]. This

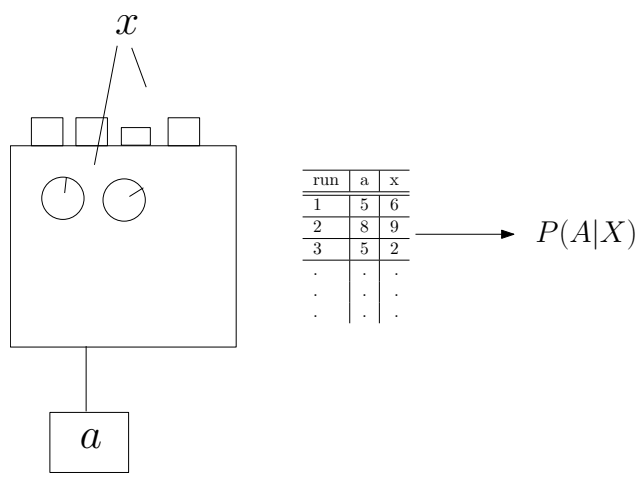

Figure 2.1: The measurement process is seen as a black box. All the configuration of knobs and buttons is encoded onto the classical variable x, which should be regarded as a label for the measurement. The output is equivalently encoded onto the classical variable a. After many repeatitions of the experiment, one can compute $P(A|X)$ just by the usual definition in terms of frequency of events.

probabilities for each input and output, one can compute the whole probability distribution. Let us denote the probability distribution by $P(A|X)$, a vector with components $P(a|x)$ for all (a, x), where $\sum_a P(a|x) = 1$, and $P(a|x) \geq 0 \ \forall (a, x)$, that is, a well-defined probability distribution.

The interesting scenarios in the device independent formalism usually involve two ore more distant observers performing measurements on its share of a physical system. Let us consider N observers, each with a measurement device. Let us denote by x_i and a_i the input and output of the i-th observer. Similarly as explained above, by collecting statistics one can compute the probability distribution $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$, a vector with components $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$. We will also use a more compact notation by defining $\vec{A} = (A_1, \ldots, A_N)$ and $\vec{X} = (X_1, \ldots, X_N)$. The probability distribution is then referred to as $P(\vec{A}, \vec{X})$. On the other hand, it is often the case that only few parties are involved. In this case they are commonly referred to as Alice, Bob, Charlie, etcetera. The probability distribution is then denoted as $P(A, B, C|X, Y, Z)$. This will be the standard notation used in most of this thesis.

becomes important in scenarios where memory effects could be exploited by a malicious agent. In section 6, for example, we avoid using the *i.i.d* assumption. However, in less paranoid scenarios, the *i.i.d* assumption fits perfectly the expected behavior of the devices.

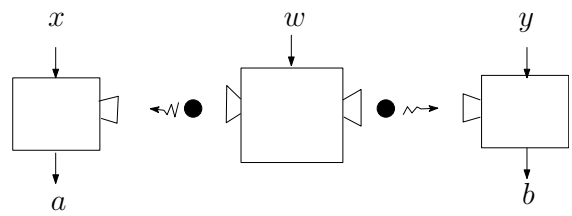

Figure 2.2: The source prepares upon request a physical system. The different preparations are labeled by the classical variable w. In this example, it is a bipartite system on which two distant observers measure x and y and obtain the outcomes a and b, respectively.

Another interesting scenario involves sources that produce physical objects that are later exposed to a measurement. Again, sources can be integrated in the device-independent formalism. This will be considered in detail in further sections. At this point, it suffices to say that the source may also be controlled by knobs that vary certain parameters. The preparation, or position of the knobs, can be encoded in the variable w. Therefore, if N observers perform measurements on a physical object prepared by the source, one can compute the probability distribution $P(A_1, \ldots, A_N | X_1, \ldots, X_N, W)$, see Fig. 2.2.

2.1.1 Assumptions

The device-independent formalism avoids as many assumption as possible about the functioning of the devices. Precisely for this reason, the assumptions that one does make play an important role and have to be well-characterized. Furthermore, as it will be explained later in detail, most of the results in the device-independent formalism can be understood as negative results such 'assumption A and B are not compatible with a certain probability distribution'. Therefore one has to characterize not only the assumptions, but how they relate to each other and which mathematical constraints impose on the probability distribution when acting together.

Measurement independence

The measurement independence assumption can be stated in the strongest version as: 'the input choice of every device is independent of the rest of the universe'. Mathematically it is expressed as $P(x|U) = P(x)$ where x

is any input involved in the experiment and U is the state of the universe. This assumption can be justified by using different arguments, that usually depend on the philosophical viewpoint of the authors, or the operational motivation of the result that is presented.

As mentioned above, the inputs represent parameters of the devices that the experimenter can vary *at will* in the laboratory. Therefore the assumption can be justified in the first place by the free will of the individuals performing the experiment to choose a certain measurement. Indeed, this assumption is often referred to as 'free-will assumption' by some authors [CK06]. The personal viewpoint of the author of this thesis is that free will is rather an ill-defined concept within a physical framework, therefore the use of this terminology is avoided.

It is indeed sufficient to assume that the input choice does not depend on the fraction of the universe that is relevant to the experiment, that is $P(x|E, U-E) = P(x|U-E)$, where E represents the state of the physical entities involved in the experiment, and $U-E$ the rest of the universe that is assumed to play no role in the results of the measurements. For instance, imagine that the inputs are chosen by tossing a coin, the measurement independence assumption just states that the coin is independent of photons, apparatuses, atoms or anything involved in the experiment, regardless of whether the coin enjoys free will. However, the situation is more intricate when using the device-independent formalism to perform tasks such as randomness generation or cryptography. To obtain acceptable generation rates of random numbers or secret bits one cannot rely on coins or experimenter choices. The input choice is integrated in the device and can be manipulated by a malicious party. All these issues are discussed deeply in section 6. In the following, we always work under the measurement independence assumption, unless the contrary is explicitly mentioned.

No-signaling

This assumption states that the choice of a measurement device cannot influence the statistics of other distant observers. It can be mathematically expressed as

$$\begin{aligned} &\sum_{a_i} P(a_1, \ldots, a_i, \ldots, a_N | x_1, \ldots, x_i, \ldots, x_N) \\ =&\sum_{a_i} P(a_1, \ldots, a_i, \ldots, a_N | x_1, \ldots, x_i', \ldots, x_N) \end{aligned} \tag{2.1}$$

for all $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N, x_1, \ldots, x_N, x_i'$ and for all $i = 1, \ldots, N$. This assumption is often justified by the laws of special relativity. If the measurements performed by the observers define space-like separated events, then the laws of special relativity prohibit that inequality (2.1) is violated. Otherwise, information about the input x_i would travel to distant observers faster than light.

Validity of quantum theory

This assumption states that the measurements processes are compatible with the laws of quantum mechanics. In particular, that exist a quantum state and quantum measurements that reproduce the statistics according to the Born rule. Mathematically it can be expressed as

$$P(a_1, \ldots, a_N | x_1, \ldots, x_N) = \mathrm{Tr}(\rho M_{a_1}^{x_1} \otimes \ldots \otimes M_{a_N}^{x_N}) \tag{2.2}$$

where ρ is a semi-definite positive operator of unit trace acting on a Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_N$. The measurement operators $M_{a_i}^{x_i}$ are semi-definite positive operators acting on $\mathcal{H}_i$, fulfilling $\sum_{a_i} M_{a_i}^{x_i} = \mathbb{I}$ for all x_i and all $i \in \{1, \ldots, N\}$. These conditions guarantee that the probability distribution is normalized and positive.

Validity of local hidden-variable models

This assumption states that the measurement processes are compatible with the laws of classical mechanics. More precisely, a probability distribution is said to be described by a local hidden-variable model (LHVM) if it can be written as

$$P(a_1, \ldots, a_N | x_1, \ldots, x_N) = \int \mathrm{d}\lambda \, p(\lambda) P_{A_1}(a_1 | x_1, \lambda) \ldots P_{A_N}(a_N | x_N, \lambda) \tag{2.3}$$

where p and P_{A_i} are well-defined probability distributions [Bel64]. This formula has a well-defined operational interpretation: the state of the whole experiment at the moment of performing the measurement is described by λ. To describe the statistics one averages over all the possible states of the experiment according to $p(\lambda)$. All the experimental devices produce the output a_i according to the information locally available to them: the input x_i and the description of the experiment λ. Apart from the hidden variable λ, there is no other correlation among the distant devices, therefore the total probability is the product of the local probability distributions, P_{A_i}.

Nonetheless, the nomenclature associated to the LHVM assumption is inspired by different viewpoints. On the one hand, the hidden variable λ can be understood as a label assigned to the state of the experiment at the moment of measuring. It is not necessary at all to assume that the state is classical nor that λ encodes all the relevant information. Therefore it is accurate to say that the only assumption therein is locality. Namely, that the outputs are generated upon the local information available to the observers: λ and the input. On the other hand, it has been shown by [Fin82, Hal09] that (2.3) is equivalent to

$$P(a_1, \ldots, a_N | x_1, \ldots, x_N) = \int \mathrm{d}\lambda \, p(\lambda) \delta_{a_1}^{A_1(\lambda, x_1)} \ldots \delta_{a_N}^{A_N(\lambda, x_N)} \tag{2.4}$$

where $A_i(\lambda, x_i) \in \{1, \ldots, d\}$. That is, the local response of the devices P_{A_i} can always be considered deterministic, or in other words, λ encodes all the relevant information of the experiment, as it can be used to predict with certainty the outcomes. Therefore, LHVM assumption is often referred to as 'local determinism' or 'local realism'.

2.1.2 Sets of probability distributions

As explained above, each of these three assumptions relates to a theoretical framework. Furthermore, each assumption defines a set of probability distributions that are compatible with the assumption. The set of all probability distributions fulfilling (2.1) will be denoted by $\mathcal{P}$. This set contains the statistics allowed by special relativity. Equivalently we denote by $\mathcal{Q}$ and $\mathcal{L}$ the set of all quantum and local correlations, respectively.

Polytopes

Let us fix some mathematical concepts that will appear recurrently, in particular, the notion of polytope (see also [BV] for more details on polytopes and convex optimization problems that may appear along this thesis).

Definition 2.1. *A polytope $\mathcal{T}$ is defined by*

$$\mathcal{T} = \{v \in \mathbb{R}^n | F_i \cdot v \leq f_i; E_j \cdot v = e_j; \; i = 1, \ldots, r \; j = 1, \ldots, s\} \tag{2.5}$$

where $F_i, E_j \in \mathbb{R}^n$ and $f_i, e_j \in \mathbb{R}$. The half-planes $F_i \cdot v \leq f_i$ are referred to as facets of the polytope $\mathcal{T}$.

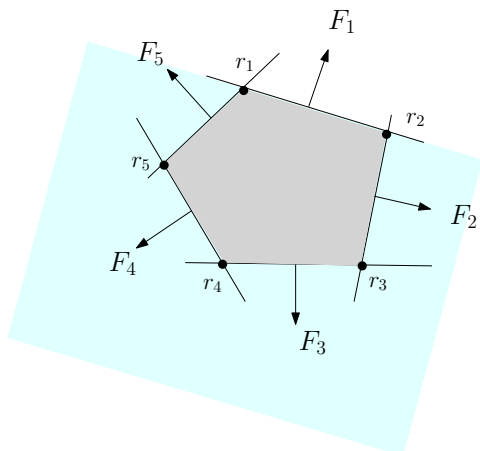

Figure 2.3: A polytope $\mathcal{T} \subset \mathbb{R}^2$, grey region. The polytope is equivalently defined either by the facets F_i (represented by normal vectors) or by the set of extremal points r_i. For instance, the condition $F_1 \cdot v \leq f_1$ would define a semi-plane (blue region). The intersection of all these semi-planes is the polytope $\mathcal{T}$

Equivalently, see Fig. 2.3 the polytope $\mathcal{T}$ can be described by the so-called extremal points. That is,

$$\mathcal{T} = \{v \in \mathbb{R}^n | v = p_1 r_1 + \ldots + p_d r_d; \sum_i p_i = 1, p_i \geq 0 \; \forall i\} \tag{2.6}$$

where $r_i \in \mathbb{R}^n \; \forall i$ are the extremal points.

Three theories, three sets

The no-signaling set $\mathcal{P}$ is defined by positivity, normalization and condition (2.1) [MAG06]. It is a convex set, as one can easily check that

$$P(\vec{A}|\vec{X}) = \sum_i p_i \, P_i(\vec{A}|\vec{X}) \tag{2.7}$$

with $p_i \geq 0 \; \forall i$ and $\sum_i p_i = 1$, is such that if $P_i(\vec{A}|\vec{X}) \in \mathcal{P} \; \forall i$, then $P(\vec{A}|\vec{X}) \in \mathcal{P}$ also. The set $\mathcal{P}$ is defined by a finite number of linear constraints, therefore it is a polytope (a polygone in higher dimensional vector space). It can be described equivalently by the set of extremal points of the polytope. The extremal points are referred to a extremal non-signaling boxes [BP05].

The quantum set $\mathcal{Q}$ is defined by condition (2.2). It is also a convex set, however it is not a polytope because the number of extremal points is not

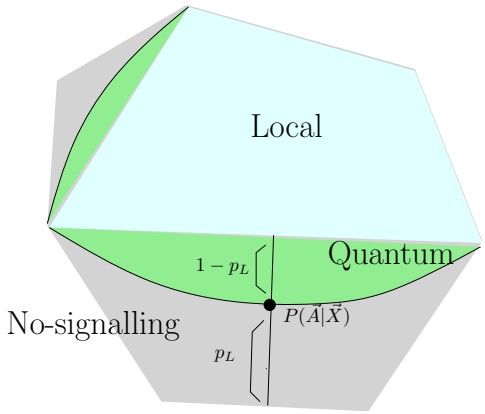

Figure 2.4: There is a strict inclusion among the three sets, so that the local set (blue) is a polytope, strictly contained into the quantum set (green). The latter is strictly contained into the no-signaling polytope (grey). A probability distribution $P(\vec{A}|\vec{X})$ is represented by the black point. It can be decomposed as a mixture of local and no-signaling probability distribution. Geometrically, the local content p_L is the distance to the local polytope

finite. One can trivially check that condition (2.2) implies (2.1), therefore $\mathcal{Q} \subseteq \mathcal{P}$. Indeed, as shown in [PR94] there exist probability distributions that lie outside the quantum set, however are non-signaling, hence $\mathcal{Q} \subset \mathcal{P}$.

The local set $\mathcal{L}$ is also polytope. According to equation (2.4) every probability distribution with a LHVM can be written as a convex combination of deterministic strategies. These are precisely the extremal points of the set $\mathcal{L}$. This set can be equivalently characterized by its facets. The local set is contained in the quantum set, and therefore in the non-signaling set, see Fig. 2.4 This can be easily seen by noticing that the hidden variable λ in equation (2.3) may be itself a quantum state $\rho_\lambda = |\lambda\rangle\langle\lambda| \otimes \ldots \otimes |\lambda\rangle\langle\lambda|$, and one can always choose quantum measurement operators such that $P(a|x,\lambda) = \mathrm{Tr}(|\lambda\rangle\langle\lambda| M_a^x)$. More surprisingly, the set $\mathcal{L}$ is strictly contained in $\mathcal{Q}$. This has been proven by Bell in his seminal work of [Bel64]. Such statement deserves development in a section of its own.

2.2 Nonlocality

The field on nonlocality relies on an apparently innocent and rather mathematical concept, $\mathcal{L} \subsetneq \mathcal{Q}$. However the implications are still nowadays a field of research in foundations of physics and recently have become a source of new applications in quantum information theory. The field of nonlocality

explores many questions related to this phenomenon: How can we quantify nonlocality?, what other properties can we infer from a probability distribution being nonlocal?, which tasks can we perform with nonlocal resources?, etcetera. In this section we review the modern formulation of Bell's theorem and many of the mathematical and conceptual tools that are used in this thesis.

2.2.1 Bell's theorem

Bell's theorem shows that $\mathcal{L} \subsetneq \mathcal{Q}$. The argument uses the so-called Bell inequalities.

Definition 2.2. *Given a vector C with real entries $C^{x_1,\ldots,x_N}_{a_1,\ldots,a_N}$, we say that*

$$C \cdot P(\vec{A}|\vec{X}) = \sum_{\substack{a_1,\ldots,a_N \\ x_1,\ldots,x_N}} C^{x_1,\ldots,x_N}_{a_1,\ldots,a_N} \, P(a_1,\ldots,a_N|x_1,\ldots,x_N) \tag{2.8}$$

is a Bell inequality if: (i) $C \cdot P(\vec{A}|\vec{X}) \leq C_{\mathcal{L}}$ where $C_{\mathcal{L}}$ is a real constant, for all $P(\vec{A}|\vec{X}) \in \mathcal{L}$; (ii) there exists another probability distribution $\tilde{P}(\vec{A}|\vec{X}) \in \mathcal{P}$ such that $C \cdot \tilde{P}(\vec{A}|\vec{X}) > C_{\mathcal{L}}$.

There exist Bell inequalities that are violated by quantum correlations, that is, such that there exists a quantum probability distribution $\tilde{P}(\vec{A}|\vec{X}) \in \mathcal{Q}$ such that $C \cdot \tilde{P}(\vec{A}|\vec{X}) > C_{\mathcal{L}}$. In fact, most Bell inequalities display a quantum violation. These violations imply that $\mathcal{L} \subsetneq \mathcal{Q}$.

Tight Bell inequality. A Bell inequality is tight whenever it corresponds to a facet of the local polytope. Tight Bell inequalities are optimal to detect whether a probability distribution belongs to the local set. To see this, consider a probability distribution $\tilde{P}(\vec{A}|\vec{X})$, that does not belong to $\mathcal{L}$. As the facets of $\mathcal{L}$ completely characterize the set, $\tilde{P}(\vec{A}|\vec{X})$ has to violate at least one of the inequalities defining the facets.

2.2.2 Nonlocality quantifiers: the local content

Whereas the violation of a Bell inequality implies nonlocality, it does not quantify it. A first attempt would be to quantify nonlocality by the amount by which a Bell inequality is violated [LVB11]. For example, take two probability distributions $P(\vec{A}|\vec{X})$ and $Q(\vec{A}|\vec{X})$ that violate a Bell inequality defined by the vector C. One is tempted to affirm that $C \cdot P(\vec{A}|\vec{X}) > C \cdot Q(\vec{A}|\vec{X})$ imply that the former is more nonlocal than the latter. However this depends crucially on the specific Bell inequality considered, as there may be

another Bell inequality for which the relation is inverted, and Q provides a larger violation than P. A more natural measurement of nonlocality can be given in terms of communication complexity [BCT99]. That is, how many bits of classical communication need the observers to exchange among them to reproduce some given nonlocal correlations. While operational appealing, this quantification has the problem that there is no algorithm to find the optimal communication protocol to reproduce some set of correlations, apart from some very specific scenarios. A more promising measure is the so-called local content [EPR92].

Definition 2.3. *Consider a non-signaling probability distribution $P(\vec{A}|\vec{X})$. The local content p_L of $P(\vec{A}|\vec{X})$ is defined as*

$$\begin{aligned} p_L &= \max_{P_{NS},P_L} q \\ \textit{such that} \quad & P(\vec{A}|\vec{X}) = q\, P_L(\vec{A}|\vec{X}) + (1-q)\, P_{NS}(\vec{A}|\vec{X}) \end{aligned} \tag{2.9}$$

where P_{NS} is an arbitrary non-signaling probability distribution fulfilling (2.1)*, and P_L is an arbitrary local probability distribution fulfilling* (2.3)*.*

The local content should be interpreted as the fraction of events that admit a local description P_L. If $P(\vec{A}|\vec{X})$ is local then $p_L = 1$, see Fig 2.4. On the other hand, a probability distribution is nonlocal if $p_L < 1$. A probability distribution is said to be maximally nonlocal if $p_L = 0$. As we will study in detail in Section 4.1, maximally nonlocal probability distributions are interesting both from a theoretical and applied point of view.

2.2.3 Multipartite nonlocality

Let us consider, for the sake of simplicity, a scenario with three parties, Alice, Bob and Charlie. A probability distribution $P(A,B,C|X,Y,Z)$ is said to be nonlocal if it cannot be written according to (2.3). However, it may be the case that the probability distribution has a decomposition as

$$P(a,b,c|x,y,z) = \sum_{\lambda} p(\lambda) P_A(a|x,\lambda) P_{BC}(b,c|y,z,\lambda). \tag{2.10}$$

If this is the case, one says that the probability distribution is local along the bipartition $A|BC$. Such probability distributions, however nonlocal, can be reproduced by just two parties acting together and, therefore, do not represent any intrinsic form of among more than two parties [Sve87]. Therefore, a stronger notion of nonlocality is needed for multipartite scenarios.

Definition 2.4. *A tripartite probability distribution $P(A,B,C|X,Y,Z)$ is said to be genuine multipartite nonlocal if it cannot be written as*

$$\begin{aligned} P(A,B,C|X,Y,Z) &= q_{A|BC}\, P_{A|BC}(A,B,C|X,Y,Z) \\ &+ q_{B|AC}\, P_{B|AC}(A,B,C|X,Y,Z) \\ &+ q_{C|AB}\, P_{C|AB}(A,B,C|X,Y,Z) \end{aligned} \tag{2.11}$$

with $q_{A|BC}+q_{B|AC}+q_{C|AB}=1$, $q_{A|BC}, q_{B|AC}, q_{C|AB} \geq 0$, and $P_{A_1|A_2A_3}$ being a probability distribution bi-local along the bipartition $A_1|A_2A_3$, (i.e. $P_{A_1|A_2A_3} = \sum_\lambda p(\lambda) P_{A_1} P_{A_2A_3}$).

Probability distributions with genuine multipartite nonlocality are discusses in deep in Section 4.2. There, we study which quantum states provide such a form of nonlocality and how it can be used to perform information tasks that are impossible in a scenario with two parties.

2.2.4 An example: The Clauser-Horne-Shimony-Holt inequality

Consider a scenario with two distant observers, Alice and Bob, that perform measurements on their share of a physical system. Both can choose between two dichotomic measurements, that is, $N = m = d = 2$. Let us label the inputs and outputs as $x, y \in \{0,1\}$ and $a, b \in \{1,-1\}$. If the results of the experiment are to be described by a LHVM, then the probability distribution can be written as

$$P(a,b|x,y) = \sum_\lambda p(\lambda) P_A(a|x,\lambda) P_B(b|y,\lambda) \tag{2.12}$$

where, without loss of generality $P_A(a|x,\lambda)$ and $P_B(b|y,\lambda)$ take values in the set $\{0,1\}$, that is, they are deterministic. Consider now, the so-called correlator defined as

$$AB_{xy} = \sum_{a,b} ab\, P(a,b|x,y) \tag{2.13}$$

One can easily check, for example by exploring all the deterministic functions P_A and P_B, that

$$C \cdot P(A,B|X,Y) \equiv AB_{00} + AB_{01} + AB_{10} - AB_{11} \leq 2 \tag{2.14}$$

is fulfilled. This inequality is known as the CHSH inequality, for Clauser-Horne-Shimony-Holt [CHSH69].

Let us study the predictions of quantum mechanics for such experiment. We will show a quantum state and quantum measurements such that the predicted probability distribution does not fulfill (2.14), therefore one can conclude that it does not admit a description in terms of LHVM.

Theorem 2.5. *[CHSH69] There exist a quantum state and a set of measurements such that the probability distribution predicted by quantum mechanics $P_Q(A, B|X, Y)$, violates the inequality* (2.14) *and therefore does not admit a LHVM description.*

Proof. Consider the singlet two-qubit quantum state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ shared by Alice and Bob. Both perform two distinct measurements on their share of the quantum state. The measurements for Alice (Bob) are described in the Pauli basis by $\hat{A}_x = \vec{A}_x \cdot \vec{\hat{\sigma}}$ ($\hat{B}_y = \vec{B}_y \cdot \vec{\hat{\sigma}}$), where $\vec{\hat{\sigma}}$ is a vector containing as entries the three Pauli matrices and $\vec{A}_x, \vec{B}_y$ are normalized vectors in $\mathbb{R}^3$, as usual in the literature. One can easily check that when measuring on the singlet state on obtains $AB_{xy} = -\vec{A}_x \cdot \vec{B}_y$. Hence, for this quantum realization $C \cdot P(A, B|X, Y) = -\vec{A}_0(\vec{B}_0 + \vec{B}_1) - \vec{A}_1(\vec{B}_0 - \vec{B}_1)$. By choosing $\vec{B}_y = -\frac{1}{\sqrt{2}}(\vec{A}_0 + (-1)^y \vec{A}_1)$ one finds that $C \cdot P(A, B|X, Y) = 2\sqrt{2}$, hence, it violates inequality (2.14). □

Thm. 2.5 implies that there exist quantum experiments that do not admit a local description, therefore quantum mechanics is said to be a nonlocal theory. Once proven that quantum mechanics can perform beyond the limits of locality, another question arises: Can quantum mechanics realize any nonlocal correlation? The answer is no. Clearly, quantum mechanics cannot be used to signal information between two distant observers, or in other words, quantum mechanics provide correlations that fulfill (2.1). The question should be sharpened: Can quantum mechanics realize any nonlocal correlation that does not lead to signaling? The answer is again no. As anticipated before $\mathcal{Q} \subset \mathcal{P}$, and this can be shown within the CHSH scenario considered here.

Theorem 2.6. *[PR94] There exist a probability distribution $P_{PR}(A, B|X, Y)$ that does not allow any signaling, that is, fulfills* (2.1)*, however it cannot be realized by quantum means.*

Proof. Consider the following probability distribution

$$P_{PR}(a,b|x,y) = \begin{cases} 1/2: & a+b = xy \mod 2 \\ 0: & \text{otherwise.} \end{cases} \tag{2.15}$$

One can easily check that $P_{PR}(A,B|X,Y)$ fulfills the non-signaling conditions. Furthermore $C \cdot P_{PR}(A,B|X,Y) = 4$. On the other hand, it is shown in [Tsi83], that $\max_{P\in\mathcal{Q}} C \cdot P(A,B|X,Y) = 2\sqrt{2}$. Therefore $P_{PR} \notin \mathcal{Q}$. □

This simple scenario allows us to detect the structure of the local, quantum and non-signaling sets. The two previous results imply that $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{P}$. Last inclusion means that quantum mechanics is not as nonlocal as the no-signaling principle allows. This can be illustrated as well by using the notion of local content of quantum correlations. Consider a decomposition of the quantum correlations used in Thm. 2.5

$$P_Q(\vec{A}|\vec{X}) = q\, P_L(\vec{A}|\vec{X}) + (1-q)\, P_{NS}(\vec{A}|\vec{X}). \tag{2.16}$$

By using linearity of Bell inequalities one can check that

$$q = \frac{C \cdot P_{NS}(\vec{A}|\vec{X}) - C \cdot P_Q(\vec{A}|\vec{X})}{C \cdot P_{NS}(\vec{A}|\vec{X}) - C \cdot P_L(\vec{A}|\vec{X})} \tag{2.17}$$

If we define $C_L \equiv \max_{P\in\mathcal{L}} C \cdot P(\vec{A}|\vec{X})$, and equivalently for C_{NS} and C_Q we obtain that the local content of $P_Q(\vec{A}|\vec{X})$.

$$p_L(P_Q) \leq \frac{C_{NS} - C_L}{C_{NS} - C_L} = 2 - \sqrt{2} \approx 0.58 \tag{2.18}$$

This implies that the Bell inequality violation provided by quantum mechanics in this scenario can be simulated by mixture of classical and non-signaling correlations. Classical correlations can be assigned a weight such that 58% of the events observed in the experiment are classical.

2.2.5 Quantum information principles

In previous sections three different sets of correlations have been characterized: the non-signaling set, the quantum set and the local set. The assumption defining the non-signaling set as a clear interpretation: no information can be transmitted arbitrarily fast. Also does the local set: the information about the system is encoded in a classical variable that determines the outcome. However, quantum correlations, despite having a clear mathematical definition in terms of Hilbert spaces, lack of an operational interpretation. It has become a field of increasing interest to find a simple statement with

an operational interpretation that describes the set of quantum correlations. Reversing the question: why some non-signaling correlations are not available in nature? The two most promising attempts are 'information causality' and 'non-triviality of communication complexity', that we review very briefly.

Information casuality. [PPK+09] It considers a scenario with two parties, Alice and Bob, that share a physical system that behaves according to the probability distribution $P(A, B|X, Y)$. Alice holds a string of n_A bits and is then allowed to send m classical bits to Bob. Information causality bounds the information Bob can gain on the n_A bits held by Alice whichever protocol they implement making use of the bipartite correlations $P(A, B|X, Y)$ and the message of m bits. It has been shown that there exist probability distributions that belong to the set $\mathcal{P}$ however violate the principle of information causality. On the other hand, all quantum probability distributions fulfill the principle. Therefore, it is allegedly a candidate to describe quantum correlations without referring to the Hilbert space structure required to formulate quantum mechanics.

Nontriviality communication complexity.[Dam05, BBL+06] Consider again a scenario with two parties, Alice and Bob, who share a probability distribution $P(A, B|X, Y)$. They are given a string of bits x_A and x_B and want to compute a function $F(x_A, x_B)$. The communication complexity of the function is the number of bits that they have to communicate in order to compute the function $F(x_A, x_B)$. It has been shown that there exist some $P(A, B|X, Y) \in \mathcal{P}$, that allow to compute any function with a constant amount of communication between the parties, in the sense that this communication is independent of the size of the vectors x_A and x_B. Such probability distributions would make communication complexity trivial. The principle of 'nontriviality communication complexity' states that correlations that make communication complexity trivial do not exist in nature.

2.2.6 Randomness

As discussed in section 2.1.1 LHVM's can be understood as models in which the outputs are generated in a deterministic causal way, see (2.4). Quantum correlations cannot be described by a LHVM, therefore one can easily anticipate that nonlocal correlations are in a sense random. This intuition has been further developed in several recent works [PAM+10, Col07], which show that nonlocality can indeed be used to generate a large number of random bits by using entangled states and a source of a few perfect random

bits.

Let us first provide a precise definition of random event. Consider a protocol generating a random variable $k \in \{0, 1\}$. Roughly speaking, one says that the variable k is a random event if k and any other classical variable e are completely uncorrelated, with e being generated by performing a measurement z on a physical system which lies outside a certain light-cone containing k. This captures the idea that any rational agent, Eve, who acquires knowledge by measuring a physical system on her possession, cannot predict the value of the variable k. Another way of looking at it is by noticing that any variable defined outside the future light-cone defined by the event k can be considered a cause of it. If k is uncorrelated with all these events, then k is a random event to which one cannot assign any cause. Apart from the independence from any potential cause, the event has to produce both outcomes with equal probablity. Putting these two things together, an ideal random bit is defined as $P_{\text{ideal}}(K, E|Z) = \frac{1}{2}P(E|Z)$, where K can take two values. Now, the randomness of an arbitrary probability distribution $P(K, E|Z)$ describing a system can then measured by any linear function of the distance between $P(K, E|Z)$ and the ideal distribution $P_{\text{ideal}}(K, E|Z)$,

$$p_{\text{guess}} = \frac{1}{2} + \frac{1}{4}\sum_{k}\max_{z}\sum_{e}|P(k, e|z) - \frac{1}{2}P(e|z)|. \tag{2.19}$$

This quantity has a well-defined operational meaning [Mas09]. Given two probability distributions $P(K, E|Z)$ and $P_{\text{ideal}}(K, E|Z) = \frac{1}{2}P(E|Z)$, p_{guess} measures the probability of successfully distinguishing which of the probability distributions describes the experiment. If $p_{\text{guess}} = \frac{1}{2}$, then $P(K, E|Z)$ is indistinguishable from a random bit. In Section 6 these ideas are further developed to define a protocol providing random bits.

2.3 Dimensionality

So far, physical tools and concepts explained above try to tackle the question: what can one say about a theory describing an experiment from the observed statistics? Clearly, this is an interesting question from a purely theoretical point of view. Furthermore, having information about the theories that are or are not capable of reproducing the experiment allows one to perform useful tasks such as cryptography and randomness amplification. However, it is easy to conceive situations in which one is interested not only in the theory describing the system, but in properties of the system within a theory. The standard procedure is to use theoretical models that one assumes to

be correct and that depend on certain parameters. The parameters that fit optimally the experimental data are the ones that one assigns to the physical system. However, this algorithm is not valid if one wants to measure the dimensionality of a physical system. It is often impossible to even conceive a model without making an assumption on the dimensionality, which is the very quantity to be measured.

The device-independent formalism can be used to tackle this question. In this thesis we develop tools to estimate the dimensionality of a physical system only from the raw statistics of an experiment. In contrast to the scenarios used in nonlocality, we do not make use of distant observers performing measurements on their share of a physical system. Instead, we consider a source of particles. This source has some tunable parameters that are encoded in the classical variable $x \in \{1, \ldots, N\}$. Then, a measurement is performed on the particle produced by the source. The measurement is described by an input $y \in \{1, \ldots, m\}$ and and output $b \in \{1, \ldots, k\}$. After repeating the experiment in order to collect reliable statistics one compute the probability distribution $P(B|X, Y)$, a vector with components $P(b|x, y)$ for all b, x, y, where $P(b|x, y)$ is the probability of obtaining output b when measurement y has been performed on preparation x. As mentioned above, the input of both the source and the measurement device is assumed to be chosen independently of the rest of the experiment. That is, the assumption of *measurement independence* is applied in the following to the variables x and y. Let us now define the sets of classical and quantum correlations that one can obtain for a fixed dimension d.

Definition 2.7. *A probability distribution $P(B|X, Y)$ admits a d-dimensional classical representation if it can be written as*

$$P(b|x, y) = \sum_{\lambda=0}^{d-1} P_S(\lambda|x) P_M(b|y, \lambda) \tag{2.20}$$

where λ is the hidden classical state of the system produced by the source according to the probability distribution $P_S(\lambda|X)$, and $P_M(b|y, \lambda)$ is the response function of the measurement device for a given hidden state λ.

Definition 2.8. *A probability distribution $P(B|X, Y)$ admits a d-dimensional quantum representation if it can be written as*

$$P(b|x, y) = Tr(\rho_x M_b^y) \tag{2.21}$$

where ρ_x is a quantum state acting on a Hilbert space of dimension d and M_b^y is a valid measurement operator such that $\sum_b M_b^y = \mathbb{I}$

In the next chapter we provide techniques to establish whether a probability distribution has d-dimensional classical or quantum models. Furthermore, we characterize the set of probability distributions with a decomposition 3.2 or 2.21.

Chapter 3

Device-independent tests for dimensionality

The scope of this chapter is to introduce techniques in the device-independent scenario that allow one to estimate the dimension of uncharacterized classical and quantum systems. The chapter is organized as follows: In section 3.1 we formalize the scenario and introduce the concept of dimension witness and discuss relevant examples. In section 3.2 we study the performance of our techniques in realistic implementations by considering the effect of detection inefficiencies. In section 3.3 we present the results obtained in an experimental realization of dimension witnesses with photons.

3.1 Device independent tests of dimensionality

In quantum mechanics, experimental observations are usually described using theoretical models which make specific assumptions on the physical system under consideration, including the size of the associated Hilbert space. The Hilbert space dimension is thus intrinsic to the model. In this chapter, the converse approach is considered: is it possible to assess the Hilbert space dimension from experimental data without an *a priori* model?

This is particularly relevant in the context of quantum information science, in which dimensionality enjoys the status of a resource for information processing. Higher dimensional systems may potentially enable the implementation of more efficient and powerful protocols. It is therefore desirable to design methods for testing the Hilbert space dimension of quantum systems which are 'device-independent'; that is, where no assumption is made on the devices used to perform the tests.

Recent years have seen the problem of testing the dimension of a non-characterized system considered from different perspectives. Initially, the concept of a dimension witness was introduced by Brunner *et al.* [BPA+08] in the context of non-local correlations. Such witnesses are essentially Bell-type inequalities, the violation of which imposes a lower bound on the Hilbert space dimension of the entangled state on which local measurements have been performed [PGWP+08, VP08, PV08, VP09, BT09, VPB10, JPPG+10]. Wehner *et al.* [WCD08] subsequently showed how the problem relates to random-access codes, and could thus exploit previously known bounds. Finally, Wolf and Perez-Garcia [WPG09] addressed the question from a dynamical viewpoint, showing how bounds on the dimensionality may be obtained from the evolution of an expectation value.

Though these works represent significant progress, they all have substantive drawbacks. The approach of Ref. [BPA+08] may not be applied to single-party systems as it is based on the non-local correlations between distant particles; the bounds of Ref. [WCD08] are based on Shannon channel capacities, which are, in general, difficult to compute; whilst the approach of Ref. [WPG09] cannot be applied to the static case. More generally, all these works show how to adapt existing techniques developed for other scenarios to the problem of assessing the dimension of a non-characterized system. However, (i) no systematic approach to this problem has yet been developed and (ii) there are no techniques specifically designed to tackle this question.

Our work bridges this gap and formalizes the problem of testing the Hilbert space dimension of arbitrary quantum systems in the simplest scenarios in which the problem is meaningful. We introduce natural tools for addressing the problem, starting by developing methods for determining the minimal dimensionality of classical systems, given certain data. Using geometrical ideas, we introduce the idea of *tight classical dimension witnesses*, leading to a generalization of quantum dimension witnesses to arbitrary systems.

3.1.1 Scenario and definitions

We consider the scenario depicted in Fig. 3.1. An initial 'black box', the *state preparator*, prepares upon requests a state—we will consider the case of both classical and quantum states. The box features N buttons which label the prepared state; when pressing button x, the box emits the state ρ_x where $x \in \{1, ..., N\}$. The prepared state is then sent to a second black box, the *measurement device*. This box performs a measurement $y \in \{1, ..., M\}$ on the state, delivering outcome $b \in \{1, ..., k\}$. The experiment is thus

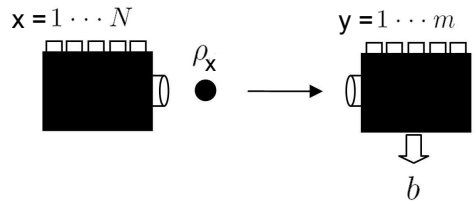

Figure 3.1: Device-independent test of classical or quantum dimensionality. Our scenario features two black boxes: a state preparator and a measurement device.

described by the probability distribution $P(B|X,Y)$, giving the probability of obtaining outcome b when measurement y is performed on the prepared state ρ_x.

Our goal is to estimate the minimal dimension of the mediating particle between the devices needed to describe the observed statistics. That is, what are the minimal classical and quantum dimensions necessary to reproduce a given probability distribution $P(B|X,Y)$?

Formally, a probability distribution $P(B|X,Y)$ admits a d-dimensional quantum representation if it can be written in the form

$$P(b|x,y) = \mathrm{tr}(\rho_x M_b^y), \tag{3.1}$$

for some state ρ_x and operators M_b^y acting on a d-dimensional Hilbert space.

We also say that probability distribution $P(B|X,Y)$ admits a d-dimensional classical representation if it can be written as

$$P(b|x,y) = \sum_{\lambda=0}^{d-1} P_S(\lambda|x) P_M(b|y,\lambda) \tag{3.2}$$

where λ is the hidden classical state of the system produced by the source according to the probability distribution $P_S(\lambda|X)$, and $P_M(b|y,\lambda)$ is the response function of the measurement device for a given hidden state λ.

This model is in the spirit of ontological models, recently investigated in Refs. [HA07, Gal09]. The classical set can be equivalently defined using quantum states. We say that $P(B|X,Y)$ admits a d-dimensional classical representation if it can be written in the form (3.1) with $[\rho_x, \rho_{x'}] = 0\ \forall x, x'$. Both definitions will be used, when convenient, along this thesis.

Definition 3.1. *The set of all probability distributions that have a decomposition of the form* (3.2) *is referred to as* $\mathcal{C}^d_{N,M,k}$. *On the other hand, the set*

of all probability distributions that can be written as convex combination of $P(B|X,Y) \in \mathcal{C}^d_{N,M,k}$ *is referred to as* $\mathcal{P}^d_{N,M,k}$*. This set is a convex polytope as it can be described by a set of extremal points.*

The distinction between these two sets will become relevant in following sections. Let us also define the concept of dimension-witness that will be used throughout this chapter.

Definition 3.2. *Given a vector* W *with real entries* $W_b^{x,y}$ *we say that*

$$W \cdot P(B|X,Y) = \sum_{b,x,y} W_b^{x,y}\ P(b|x,y) \tag{3.3}$$

is a classical dimension-witness *if (i)* $W \cdot P(B|X,Y) \leq C_d$*, where* C_d *is a real constant, for every* $P(B|X,Y)$ *with a classical description of dimension* d*, (ii) there exist another probability distribution* $\tilde{P}(B|X,Y)$ *with a classical description of dimension* $\tilde{d} > d$ *such that* $W \cdot \tilde{P}(B|X,Y) > C_d$*.*

Definition 3.3. *Given a vector* W *with real entries* $W_b^{x,y}$ *we say that*

$$W \cdot P(B|X,Y) = \sum_{b,x,y} W_b^{x,y}\ P(b|x,y) \tag{3.4}$$

is a quantum dimension-witness *if (i)* $W \cdot P(B|X,Y) \leq Q_d$*, where* Q_d *is a real constant, for every* $P(B|X,Y)$ *with a quantum description of dimension* d*, (ii) there exist another probability distribution* $\tilde{P}(B|X,Y)$ *with a quantum description of dimension* $\tilde{d} > d$ *such that* $W \cdot \tilde{P}(B|X,Y) > Q_d$*.*

3.1.2 Dimension-witnesses from the classical polytope

Tight classical dimension witnesses.

We start by deriving a general method for finding a lower bound on the dimensionality of the classical states necessary to reproduce a given probability distribution $P(B|X,Y)$. For simplicity we shall focus on measurements with binary outcomes, which we denote $b = \pm 1$; the generalization to larger alphabets is straightforward. It then becomes convenient to use expectation values:

$$E_{xy} = P(b = +1|x,y) - P(b = -1|x,y). \tag{3.5}$$

Every experiment is characterized by a vector of correlation functions

$$\vec{E} = (\vec{v}_{x=1}, \vec{v}_{x=2}, ..., \vec{v}_{x=N}), \tag{3.6}$$

where $\vec{v}_x = (E_{x1}, E_{x2}, ..., E_{xm})$ is a vector containing the correlation functions for a given preparation x and all measurements. Deterministic experiments—those in which only one outcome appears for any possible pair of preparation and measurement—correspond to vectors $\vec{E}_{\text{det}}$ for which $E_{xy} = \pm 1$ for all x, y. Clearly, any possible experiment may be written as a convex combination of deterministic vectors $\vec{E}_{\text{det}}$. Thus, the set of all possible experiments defines a polytope—*i.e.* a convex set with a finite number of extremal points—denoted by $\mathcal{P}_{N,M,2}$. The facets of $\mathcal{P}_{N,M,2}$ are termed positivity facets, of the form $E_{xy} \leq 1$ and $E_{xy} \geq -1$, which ensures that probabilities $P(b|x, y)$ are well defined. Thus $\mathcal{P}_{N,M,2}$ may be viewed as the set of all valid probability distributions. Note that $\mathcal{P}_{N,M,2}$ resides in a space of dimension NM and has 2^{NM} vertices, corresponding to the deterministic vectors $\vec{E}_{\text{det}}$.

Next, we would like to characterize the set of realizable experiments in the case that the dimension d of the classical states is limited. We first note that if $d \geq N$, all possible experiments can be realized. Indeed, it is then possible to encode the choice of preparation x in the classical state; *i.e.* $p(\lambda|x) = \delta_x^\lambda$. Thus, any probability distribution $P(B|X, Y)$—*i.e.* any vector $\vec{E}$ in $\mathcal{P}_{N,M,2}$—can be obtained, since the measurement device has full information of both x and y.

Therefore the problem of bounding the dimension of classical (or quantum) systems necessary to reproduce a given set of data is meaningful only if $d < N$. In this case, it turns out that not all possible experiments can be realized. Let us first focus on deterministic experiments. Clearly, if the classical state sent by the state preparator is of dimension $d < N$, then (at least) $\lceil N/d \rceil$ preparations must correspond to the same state (*i.e.* the same classical dit). Therefore, only a subset of the 2^{NM} deterministic vectors can be obtained in this case: those deterministic vectors $\vec{E}^d_{\text{det}}$ composed of (at least) $\lceil N/d \rceil$ vectors $\vec{v}_x$ which are the same.

General strategies consist of mixtures of these deterministic points. It is however possible to identify two different scenarios. In the first scenario, the state preparator and the measurement device share no pre-established correlations and, thus, mix different deterministic preparations and measurements in an uncorrelated manner. In a practical setup, this is often a very reasonable assumption. In this case, the set of experiments realizable with a d-dimensional classical system is $\mathcal{C}^d_{N,m,2}$. This set is not convex, as not every mixture of points $\vec{E}^d_{\text{det}}$ has a decomposition of the form (3.2). This scenario will be considered in detail in Section 3.2. In the second scenario, the state preparator and the measurement device share classical correlations.

This is the natural situation in a device-independent scenario, where no assumption about the devices is possible. Now, the set of realisable points is by construction convex and corresponds to the convex hull of deterministic vectors $\vec{E}^d_{\text{det}}$, a polytope denoted $\mathcal{P}^d_{N,M,2}$. In this section, we focus on the second scenario since: (i) its characterization is simpler, as a polytope is defined by a finite set of linear inequalities and (ii) it is more general, as any experiment in the first scenario is contained in $\mathcal{P}^d_{N,M,2}$.

The polytope $\mathcal{P}^d_{N,M,2}$ is a strict subset of $\mathcal{P}_{N,M,2}$. Thus it features additional facets which are not positivity facets. These new facets are 'tight classical dimension witnesses' (for systems of dimension d), and are formally given by linear combinations of the expectation values E_{xy}; *i.e.*

$$\vec{W} \cdot \vec{E} = \sum_{x,y} w_{xy} E_{xy} \leq C_d \tag{3.7}$$

where the probabilities (entering E_{xy}) are of the form of Eq. (3.2), a classical representation of dimension d. These inequalities are classical dimension witnesses in the sense that: (i) for any experiment involving classical states of dimension d, the associated correlation vector $\vec{E}$ will satisfy inequality (3.7); (ii) in order to violate inequality (3.7), classical systems of dimension strictly larger than d are required. Note that a witness is termed 'tight' when it corresponds to a facet of the polytope $\mathcal{P}^d_{N,M,2}$; this terminology is borrowed from the study of non-locality, in analogy to tight Bell inequalities.

To summarize, by characterizing the polytopes $\mathcal{P}^d_{N,M,2}$ (that is, by finding all the facets of $\mathcal{P}^d_{N,M,2}$) one can lower bound the dimension of the classical systems necessary to reproduce a given probability distribution $P(B|X,Y)$. Clearly, if a probability distribution is proven not to belong to $\mathcal{P}^d_{N,M,2}$, it requires classical systems of dimension strictly larger than d. In the case that the state preparator and the measuring device are allowed to share pre-established correlations, our technique also provides an upper bound on the dimension, since all experiments in $\mathcal{P}^d_{N,M,2}$ can then be obtained from classical systems of dimension d. In this case our methods makes it possible, in principle, to determine the minimum dimensionality required in order to reproduce any given probability distribution.

Quantum dimension witnesses

The above ideas can be extended to the problem of finding lower bounds on the Hilbert space dimension of quantum systems necessary to reproduce a certain probability distribution. We first define linear quantum d-

dimensional witnesses as linear expression of the form

$$\vec{W} \cdot \vec{E} = \sum_{x,y} w_{xy} E_{xy} \leq Q_d, \tag{3.8}$$

where the correlation functions E_{xy} can be written in terms of probabilities of the form (3.1) with ρ_x acting on $\mathbb{C}^d$, and there exists a probability distribution $P(B|X,Y)$ such that $\vec{W} \cdot \vec{E} > Q_d$. This generalises the concept of dimension witness of Ref. [BPA+08] to arbitrary quantum systems.

It would be, in general, very interesting to fully characterize the set of experiments, *i.e.* of vectors $\vec{E}$, that can be obtained from quantum states of a given dimension. Indeed, this would allow one to determine the minimal Hilbert space dimension necessary to reproduce any given probability distribution. As above, it is possible to define different scenarios, depending on whether the state preparator and the measurement device share correlations, which can now be quantum. In the case of no correlations, the set of realizable points is again not convex. In the case of correlated devices, the set of quantum experiments is convex. However, obtaining its complete characterization represents a more difficult problem, since it is not a polytope. That is, the number of extreme points is infinite and its boundary cannot be characterized by a finite number of linear dimension witnesses. All these different scenarios will be discussed in Section 3.2. As stated, for the sake of simplicity, our analysis here is restricted to devices sharing classical correlations.

3.1.3 Case studies

As an application of our general formalism, we now present several examples of dimension witnesses. In particular, we give a family of linear witnesses which can be used as both a classical and quantum witness for any dimension. In general, the classical and quantum bounds of our witnesses—C_d and Q_d, respectively—differ, and thus our witnesses can distinguish between classical and quantum resources of given dimensions. We also give an example of a non-linear witness for qubits.

Simplest case

We start by considering the case $d = 2$, *i.e.* where the classical state sent by the state preparator is simply a bit. Indeed, we saw above that our problem is meaningful only if $d < N$, and thus we consider the case of three preparations ($N = 3$) and two measurements ($M = 2$) with binary

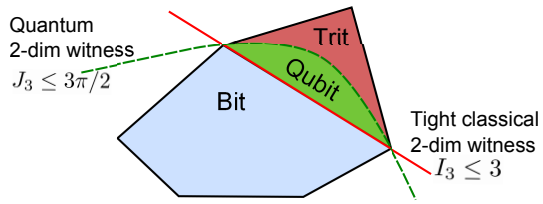

Figure 3.2: Schematic representation of the sets of experiments achievable from classical and quantum states of given dimensions for case study 1. The set of experiments (more precisely its convex hull) attainable from 2-dimensional classical states, *i.e.* bits, forms the polytope $\mathcal{P}^2_{3,2,2}$ (blue region). The inequality $I_3 \leq 3$ (solid line), a facet of this polytope, is a 'tight 2-dimensional classical witness'. The set of experiments attainable from 2-dimensional quantum states, *i.e.* qubits, (green and blue region) is strictly larger. The inequality $J_3 \leq \frac{3\pi}{2}$ (dashed curve) is a qubit-witness; it cannot be violated by performing measurements on qubits: qutrits are required. The set of all possible experiments (blue, green and red regions) forms the polytope $\mathcal{P}_{3,2,2}$; any point in it can be reproduced with a trit or a qutrit.

outcomes [1]. We fully characterize the polytope $\mathcal{P}^2_{3,2,2}$. It features a single type of non-trivial facet given by

$$I_3 \equiv |E_{11} + E_{12} + E_{21} - E_{22} - E_{31}| \leq 3. \tag{3.9}$$

This inequality is a tight 2-dimensional classical witness. To be violated, trits (or higher-dimensional systems) are required. Note that trits are sufficient to reach the algebraic maximum of $I_3 = 5$; indeed any correlation vector $\vec{E}$ in $\mathcal{P}_{3,2,2}$ can be obtained using trits. Fig. 3.2 shows a schematic view of the situation.

The witness I_3 is also a 2-dimensional quantum witness. The maximal value of I_3 obtainable from qubits can be computed analytically. Here the analysis may be restricted to pure states, since I_3 is a linear expression of the probabilities, and to rank-one projective measurements, since we consider measurements of two outcomes [Mas05]. By solving the maximization problem, it can be shown that $\max_{\rho \in \mathcal{B}(\mathbb{C}^2)} I_3 = 1 + 2\sqrt{2} \approx 3.8284$. The first four terms in Eq. (3.9) can be seen as the CHSH polynomial, whose maximum quantum value is equal to $2\sqrt{2}$. This maximization does not involve the third preparation, which can be always chosen such that

[1] Note that the CHSH polynomial does not work as a dimension witness, since it features only two preparations—indeed, it is necessary to have $d < N$. In the device-independent scenario considered here, it is possible to reach the maximum of CHSH=4 by sending a classical bit (the bit simply indicates which preparation has been chosen).

$E_{31} = -1$. In order to quantum mechanically reproduce a probability distribution $P(B|X,Y)$ leading to $I_3 > 1 + 2\sqrt{2}$, qutrits (or systems of higher dimension) are required; in fact classical trits would suffice. The maximal qubit value can be obtained from the following preparations and measurements: $\rho_x = (\mathbb{I} + \vec{r_x} \cdot \vec{\sigma})/2$, $M_b^y = (\mathbb{I} + b\vec{s_y} \cdot \vec{\sigma})/2$ with $\vec{s}_1 = (\vec{r}_1 + \vec{r}_2)/\sqrt{2}$, $\vec{s}_2 = (\vec{r}_1 - \vec{r}_2)/\sqrt{2}$, $\vec{r}_3 = (-\vec{r}_1 - \vec{r}_2)/\sqrt{2}$, and where $\vec{\sigma} = \{\sigma_x, \sigma_y, \sigma_z\}$ denotes the vector of Pauli matrices. Indeed, the correlation functions are then simply given by $E_{xy} = \vec{r_x} \cdot \vec{s_y}$.

An interesting feature of the witness I_3 is that it can also distinguish between classical and quantum resources of a given dimension; here, bits and qubits. If the inequality (3.9) (or one of its symmetries) is violated by a given probability distribution, then it follows that qubits, rather than classical bits, have been used. It is interesting to contrast this result with the Holevo bound [Hol73], which shows that one qubit cannot be used to send more than one bit of information. In our scenario, the state of the mediating particle somehow encodes the information about the value of the classical value x. However, here the use of quantum particles does provide an advantage.

Furthermore, we have strong numerical evidence that the following inequality (based on I_3) is never violated by qubits:

$$\begin{aligned} J_3 \equiv |\arcsin E_{11} + \arcsin E_{12} + \arcsin E_{21} \\ - \arcsin E_{22} - \arcsin E_{31}| \leq \frac{3\pi}{2}, \end{aligned} \tag{3.10}$$

suggesting that J_3 may be used as a non-linear dimension witness. Moreover, the bound is tight, in the sense that there exist qubit preparations and measurements attaining it—for instance the states and measurements leading to $I_3 = 1 + 2\sqrt{2}$ given above.

Generalization.

Next we generalize the witness I_3 presented above, in order to obtain classical and quantum dimension witnesses for any dimension. The form of I_3—see Eq. (3.9)—suggests the following natural generalization for the case $N = M + 1$:

$$I_N \equiv \sum_{j=1}^{N-1} E_{1j} + \sum_{i=2}^{N} \sum_{j=1}^{N+1-i} \alpha_{ij} E_{ij} \tag{3.11}$$

$$\text{with } \alpha_{ij} = \begin{cases} +1 & \text{if } i + j \leq N, \\ -1 & \text{if } i + j = N + 1. \end{cases}$$

	C_2 (bit)	Q_2 (qubit)	C_3 (trit)	Q_3 (qutrit)	C_4 (quat)
I_3	3	$1+2\sqrt{2}$	5	5	5
I_4	5	6	7	7.9689	9

Table 3.1: Classical and quantum bounds for the dimension witnesses I_3 and I_4. Notably, these witnesses can distinguish classical and quantum systems of given dimensions.

It can be erified that for classical states of dimension $d \leq N$, the following relation holds:

$$I_N \leq L_d = \frac{N(N-3)}{2} + 2d - 1. \tag{3.12}$$

Indeed for $d = N$ one obtains the algebraic bound $I_N = L_{d=N} = N(N+1)/2-1$. Using the methods of Ref. [Mas02] we have checked that the inequality $I_N \leq L_{d=N-1}$ is a tight classical dimension witnesses (*i.e.* a facet of the polytope $\mathcal{P}^d_{N,M,2}$ with $M = d = N-1$) for $N \leq 5$. Based on this evidence, we conjecture that it is a tight witness for all values of N.

Next we show that the inequality $I_N < L_{d=N}$ is a quantum dimension witness. More precisely, it is impossible to reach the algebraic bound of I_N by performing measurements on quantum states of dimension $d = N-1$. Since I_N is a linear expression of expectation values, it is sufficient to consider pure states, and one may write $E_{ij} = \langle\psi_i|O_j|\psi_i\rangle$, where $O_j = M^j_{+1} - M^j_{-1}$ is the measured quantum observable. Clearly, in order to reach the algebraic maximum of I_N we require $E_{ij} = \text{sign}[\alpha_{ij}]$ for $i + j \leq N+1$, and thus the states $\{|\psi_i\rangle\}$ must be eigenstates of the observables $\{O_j\}$ with eigenvalues $\{\text{sign}[\alpha_{ij}]\}$. From the structure of I_N, it can be seen that for any pair of preparations $|\psi_s\rangle$ and $|\psi_t\rangle$ with $1 \leq s < t \leq N$, the observable O_{N-t+1} must have eigenvalue $+1$ for $|\psi_s\rangle$ and eigenvalue -1 for $|\psi_t\rangle$. Thus all preparations must be mutually orthogonal, since any pair of states $|\psi_s\rangle$ and $|\psi_t\rangle$ can be perfectly distinguished by measuring observable O_{N-t+1}. Since we must consider N mutually orthogonal preparations, a Hilbert space of dimension (at least) $d = N$ is required to reach the algebraic maximum of I_N. It therefore follows that the inequality $I_N < L_{d=N}$ is a dimension witness for quantum systems of dimension $d = N-1$.

We believe, however, that better bounds can be obtained for the expression I_N. This is the case for $N = 3$, as shown above, as well as for $N = 4$ where we have been able to compute numerically the bounds for qubits and qutrits. These results are summarized in Table 1. Indeed, it would be desirable to find tight bounds for the witness I_N for quantum states of any Hilbert space dimension $d < N$.

3.1.4 Conlusions

We have addressed the problem of testing the dimensionality of classical and quantum systems in a device-independent scenario. We have introduced the concept of 'tight classical dimension witnesses' which allows one to put a lower bound on the dimensionality of classical states necessary to reproduce certain data. This naturally led us to generalize the concept of quantum dimension witnesses to arbitrary quantum systems. To illustrate these ideas, we have provided explicit examples of dimension witnesses. We have shown that these witnesses (i) are tight for small number of classical preparations, (ii) work both as classical and as quantum dimension witnesses, and (iii) allow one to distinguish classical and quantum states of given dimensions. Finally, we have introduced non-linear dimension witnesses, and have presented an example of such a witness for the simplest scenario.

3.2 Detection Loophole in dimension witnesses

Any experimental implementation of dimension witnesses is unavoidably affected by noise and loss, thus making important the question whether it is possible to perform reliable dimension witnessing with non-optimal detection efficiency. Despite its relevance for experimental implementations and practical applications, this problem is usually referred to as detection loophole, and it is equivalently present in the experimental violation of Bell inequalities [CH74]. In this section we study the performance of dimension witnesses under detection inefficiencies. We calculate bounds on the minimum detection efficiency required to bound the dimension. We show that the detection loophole imposes a serious limitation when measuring systems of high dimensionality. However, we overcome this negative result by adding a mild assumption on the devices, namely, that source and measurement device do not share correlations. In this case, we exploit the non convexity of the resulting set of correlations to show that the detection loophole can be closed for any (non-null) value of the detection efficiency.

3.2.1 Minimum efficiencies

Consider a scenario as in Fig 3.1. If the measurement device has detection inefficiencies, then one registers events in which none of the outcomes $b \in \{1, ..., k\}$ is observed. This events are called no-click events and can be integrated as a new outcome in the formalism. Therefore, in an imperfect experiment one observes a probability distribution of $k+1$ outcomes, where the last outcome corresponds to the no-click event. This probability distribution is denoted by $P^\eta(B|X,Y)$ and we model an arbitrary device subjected to imperfections as

$$P^\eta(b|x,y) = \mathrm{tr}(\rho_x \Pi_b^y), \tag{3.13}$$

with $\Pi_b^y = \eta M_b^y, \quad \forall b \in \{1, ..., k\}$ and $\Pi_{k+1}^y = (1-\eta)\mathbb{I}$, where M are valid measurement operators with $\sum_{b=1}^k M_b^y = \mathbb{I}$. That is, we are assuming that an ideal quantum realization with measurement operators M_b^y is affected by inefficiencies in such a way that the no-click event occurs with probability $(1-\eta)$. We make here the implicit assumption that the probability of a no-click event does not depend neither on the preparation x, nor on the measurement y. This is a natural assumption if the inefficiencies are due to losses in optical fibers or no-clicks on photo-detectors [2].

[2] It is worth to point out that these assumptions do not compromise the device-independent approach. We are making assumptions about how inefficiencies affect the

We note that by definition (3.13)

$$P^{\eta}(B|X,Y) = \eta P^{\mathrm{p}}(B|X,Y) + (1-\eta)P^{\mathrm{nc}}(B|X,Y) \tag{3.14}$$

where $P^{\mathrm{p}}(b|x,y) = \mathrm{Tr}(\rho_x M_b^y)$ for $b \in \{1,\ldots,k\}$ and $P^{\mathrm{p}}(k+1|x,y) = 0$, that is, $P^{\mathrm{p}}(B|X,Y)$ is the probability distribution that one would obtain in a perfect experiment without inefficiencies; and $P^{\mathrm{nc}}(b|x,y) = \delta^b_{k+1}$, that is, a probability distribution that always output no-click. Let us define the two parameters that we investigate

Definition 3.4. *Given a quantum dimension witness W such $W \cdot P(B|X,Y) \leq Q_d$ for all $P(B|X,Y)$ with a d-dimensional description, we define η_{dim} as the minimum value of η such $W \cdot P^{\eta}(B|X,Y) > Q_d$ while $P^p(B|X,Y)$ has a $d+1$-dimensional realization. In other words, consider a $d+1$-dimensional system affected by inefficiencies, then η_{dim} quantifies the minimum efficiency required to violate a d-dimensional dimension witness.*

Definition 3.5. *Given a dimension witness W such $W \cdot P(B|X,Y) \leq C_d$ for all $P(B|X,Y)$ with a d-dimensional classical description, we define η_{qc} as the minimum value of η such $W \cdot P^{\eta}(B|X,Y) > C_d$ while $P^p(B|X,Y)$ has a d-dimensional* quantum *realization. In other words, consider a d-dimensional quantum system affected by inefficiencies, then η_{qc} quantifies the minimum efficiency required to violate a d-dimensional classical dimension witness.*

In general, η_{dim} and η_{qc} depend on the particular dimension witness considered. We will focus on the class of dimension witnesses considered in (3.23), with $N+1 = d$, thus we denote $W_{d+1} \cdot P(B|X,Y) \equiv I_{d+1}$

By linearity of dimension-witnesses and normalization of the probability distribution one can easily check that

$$W_{d+1} \cdot P^{\eta}(B|X,Y) = \eta\, W_{d+1} \cdot P_{\mathrm{p}}(B|X,Y). \tag{3.15}$$

and by using the bounds imposed by (3.12) and the fact that $\max_{P(B|X,Y)} W_{d+1} \cdot P(B|XY) \geq \max_{P(B|X,Y)} W_d \cdot P(B|XY) + 1$ (see [DAPGA12] for details) we obtain that

$$\frac{d-1}{d} \leq \eta_{\mathrm{qc}} \leq \frac{d-1}{d-2+\sqrt{2}} \tag{3.16}$$

probability distribution designed to violate a dimension-witness. However, once the probability distribution violates a dimension-witness one can lower bound the dimension regardless of the validity of those assumptions.

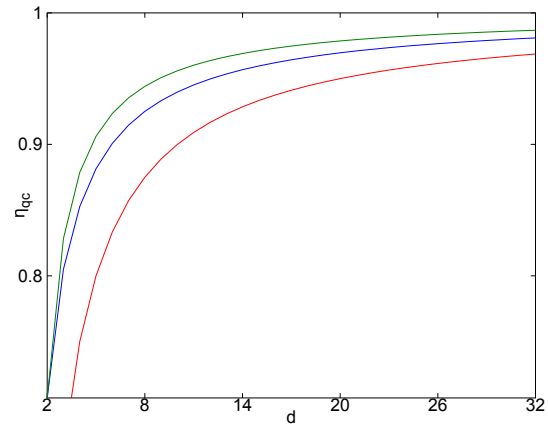

Figure 3.3: (Color on line) Threshold value (middle line) of the detection efficiency $\eta_{\rm qc}$ obtained after numerical optimization as a function of the dimension d. Lower bound (lower line) and upper bound (upper line) given by Eq. (3.16) are also plotted. Upper bound is tight for $d = 2$. The detection efficiency $\eta_{\rm qc}$ asymptotically goes to 1 as $d \to \infty$ since its upper and lower bound do the same.

and

$$1 - \frac{2-\sqrt{2}}{d} \leq \eta_{\rm dim}. \tag{3.17}$$

These lower bounds imply that required efficiency tends to one as the dimension increases, hence the detection loophole plays a dramatic role when dealing with high-dimensional systems, see Fig. 3.3 and 3.4.

3.2.2 Closing the detection loophole with an extra assumption

Throughout the previous chapters no assumption is made about the functioning of the devices. In particular, the *preparator* and the *measurement device* are allowed to share classical correlations. For this reason, the set of correlations that admit a d-dimensional classical description is a convex polytope, that we referred to as $\mathcal{P}^d_{N,M,k}$ in a scenario of N preparations, M measurements and k outcomes. If one does not allow for shared pre-established correlations between apparatuses, then the set of d-dimensional classical correlations is not convex in general and it is not possible to characterize it by its facets. Nevertheless, these non-convex sets of correlations are interesting when considering scenarios with inefficient detectors. We show, that under the assumption that the devices do not share classical correla-

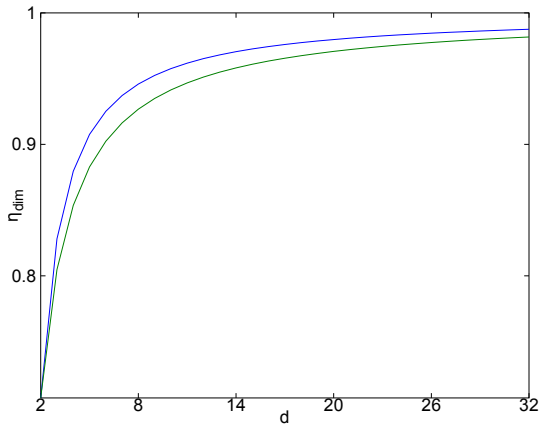

Figure 3.4: (Color on line) Threshold value (upper line) of the detection efficiency η_{dim} after numerical optimization as a function of the dimension d. Lower bound (lower line) given by Eq. (3.17) is also plotted. Lower bound is tight for $d = 2$. The detection efficiency η_{dim} asymptotically goes to 1 as $d \to \infty$ since its lower bound does the same (and $\eta_{\text{dim}} \leq 1$ is a trivial upper bound).

tions, it is possible to certify the dimension of a classical system for any non-null value of the efficiencies.

Classical sets without shared randomness.

We recall definition (3.2). The set of all probability distributions with such decomposition is denoted by $\mathcal{C}^d_{N,M,k}$. This set is to be understood as the correlations that one can perform with a d-dimensional classical system when the devices do not share pre-established correlations. If one allows the source and the measurement device to share correlations, then any convex mixture of d-dimensional classical strategies can be performed, that is, the set $\mathcal{P}^d_{N,M,k}$. Interestingly, the set $\mathcal{C}^d_{N,M,k}$ is nonconvex and thus, strictly contained in the set $\mathcal{P}^d_{N,M,k}$. An example can be found in the simplest scenario.

Lemma 3.6. *In a scenario with $N = 3, M = 2, k = 2, d = 2$, the set $\mathcal{C}^2_{3,2,2}$ is nonconvex and strictly contained in $\mathcal{P}^2_{3,2,2}$*

Proof. Consider the probability distribution $P(B|X,Y)$ defined as

$$P(B|X,1) = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix}, \qquad P(B|X,2) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \tag{3.18}$$

where, the rows indicate the value of x and columns the value of b. One can easily check that this probability distribution does not violate the dimension witness (3.9), therefore it belongs to the set $\mathcal{P}^2_{3,2,2}$. However, it does not belong to the set $\mathcal{C}^2_{3,2,2}$. Let us prove this by *reductio ad absurdum*. Suppose there exists a decomposition of the form (3.2) for the previous probability distribution. Since the measurement $y = 1$ is deterministic for $x = 1, 3$ and measurement $y = 2$ is deterministic for $x = 2$, the three preparations must be deterministic, that is $P_S(\lambda|x) \in \{0, 1\}$. As λ can only take two possible values, it must be the case that $P(b|x, y) = P(b|x', y)\ \forall\ y$ for at least two preparations x and x'. One can clearly see in tables above that this is not the case, and thus, one arrives to a contradiction. □

Detection loophole without shared correlations

We will study the performance of imperfect devices under the assumption that they do not share pre-established correlations. Consider the probability distribution affected by inefficiencies as defined in (3.14).

Theorem 3.7. *If $P^p(B|X, Y) \notin \mathcal{C}^2_{N,M,k+1}$ then $P^\eta(B|X, Y) \notin \mathcal{C}^2_{N,M,k+1}$ for any value of η.*

Proof. Let us recall first that we are assuming that the efficiencies do not depend neither on the preparation x nor on the measurement performed y. Actually, the proof holds only by assuming that the efficiency is independent of x, (see [DAPG+12] for details). That is,

$$P^\eta(b = k + 1|X, Y) = P^\eta(b = k + 1|Y). \tag{3.19}$$

Let us now prove a converse equivalent statement to the one of the theorem. That is, we show that if $P^\eta(B|X, Y) \in \mathcal{C}^2_{N,M,k+1}$ then $P^{\mathrm{p}}(B|X, Y) \in \mathcal{C}^2_{N,M,k+1}$. The former, see (3.2), implies that

$$P^\eta(b|x, y) = \sum_{\lambda=0}^{1} P^\eta_S(\lambda|x) P^\eta_M(b|y, \lambda) \tag{3.20}$$

which together with (3.19) implies that

$$[P^\eta_S(\lambda = 0|x) - P^\eta_S(\lambda = 0|x')][(P^\eta_M(b = k+1|\lambda = 0, y) - (P^\eta_M(b = k+1|\lambda = 1, y)] = 0 \tag{3.21}$$

for all x, x'. This clearly implies that at least one of the brackets has to be equal to zero. If $[P^\eta_S(\lambda = 0|x) - P^\eta_S(\lambda = 0|x')] = 0\ \forall x, x'$, then the

message λ sent by Alice to Bob does not depend on the input x, that is, $P_S^\eta(\lambda|x) = P_S^\eta(\lambda)\ \forall \lambda, x$, therefore $P_S^\eta(B|X,Y) = P_S^\eta(B|Y)$. As the efficiency is assumed to be independent of x, by (3.14), one has that $P^{\mathrm{p}}(B|X,Y) = P^{\mathrm{p}}(B|Y)$, which clearly belongs to $\mathcal{C}^2_{N,M,k+1}$ as it can be simulated without any message being send from Alice to Bob.

On the other hand, if $[(P_M^\eta(b = k+1|\lambda = 0, y) - (P_M^\eta(b = k+1|\lambda = 1, y)] = 0$, by (3.14) and (3.19) one has that

$$P^{\mathrm{p}}(b|x,y) = \sum_{\lambda=0}^{1} P_S(\lambda|x)\Big(\frac{1}{\eta}(P_M^\eta(b|\lambda,y) - (1-\eta)\delta^b_{k+1})\Big). \tag{3.22}$$

By defining $P_M^{\mathrm{p}}(b|\lambda,y) \equiv \Big(\frac{1}{\eta}(P_M^\eta(b|\lambda,y) - (1-\eta)\delta^b_{k+1})\Big)$, one obtains a valid model of the form (3.2) for $P^{\mathrm{p}}(B|X,Y)$, which implies that it belongs to $\mathcal{C}^2_{N,M,k+1}$. One can easily check that $P_M^{\mathrm{p}}(b|\lambda,y)$ is a positive and normalized probability distribution. □

This theorem has very useful implications for experimental implementations of dimension witnesses. Consider a source that it is not heralded, that is, the experimenter does not have a record of the no-click events (or cannot distinguish this event from an absence of emission from the source) and they do not enter into the statistics. Then, the computed statistics are precisely $P^{\mathrm{p}}(B|X,Y)$. If this probability distribution violates a d-dimensional dimension witness one can conclude that it does not belong to the polytope $\mathcal{P}^d_{N,M,k+1}$, hence, does not belong to the set $\mathcal{C}^d_{N,M,k+1}$, as this set is contained in the former. Then, by using Theorem 3.7 one can infer that the real probability distribution that the experimenter had no access to, $P^\eta(B|X,Y)$, does not belong to $\mathcal{C}^d_{N,M,k+1}$ either. Thus the experiment requires at least a $d+1$-dimensional description under the assumption that the devices do not share classical correlations.

3.3 Experimental implementation

The concept of dimension witness is appealing because it is specifically designed for an experimental implementation. All the theoretical machinery is distilled into a linear combination of the measurement statistics. Furthermore, the approach can be applied to any kind of physical system: classical, quantum, photons, atoms, etcetera. In particular, due to its interest in quantum information science, we have conducted an experiment on photons, manipulated on polarization and angular momentum [HGM+12]. We have shown experimentally the violation of a dimension witness and certified dimensionality up to dimension four. Furthermore, we are able to measure the distinction between classical and quantum systems of a given dimension. We note that the interest of our dimension-witness techniques has driven another group to a parallel experimental demonstration of dimension witness. This parallel demonstration has been conducted with photons manipulated on polarization and on spatial modes [ABCB12]. These two approaches highlight the versatility of the dimension-witness approach.

3.3.1 Experiment with polarization and angular momentum of photons

Here we shall focus on a dimension witness (3.23), for a scenario consisting of $N = 4$ possible preparations and $M = 3$ measurements with only two possible outcomes, labeled by $b = \pm 1$:

$$I_4 \equiv E_{11} + E_{12} + E_{13} + E_{21} + E_{22} - E_{23} + E_{31} - E_{32} - E_{41}, \tag{3.23}$$

where $E_{xy} = P(b = +1|x, y) - P(b = -1|x, y)$. The witness I_4 can distinguish ensembles of classical and quantum states of dimensions up to $d = 4$. All the relevant bounds are summarized in Table 3.1.

In order to test this witness experimentally, we must generate classical and quantum states of dimension 2 (bits and qubits, respectively), classical and quantum states of dimension 3 (trits and qutrits), and classical states of dimension 4 (quarts). To do so we exploit the angular momentum of photons [MTTT07, MVWZ01], which contains a spin contribution associated with the polarization, and an orbital contribution associated with the spatial shape of the light intensity and its phase. Within the paraxial regime, both contributions can be measured and manipulated independently. The polarization of photons is conveniently represented by a 2-dimensional Hilbert space, spanned by two orthogonal polarization states (e.g., horizontal and vertical). The spatial degree of freedom of light lives

in an infinite-dimensional Hilbert space [MTTT01], spanned by paraxial Laguerre-Gaussian (LG) modes. LG beams carry a well-defined orbital angular momentum (OAM) of $m\hbar$ (m is integer) per photon that is associated with their spiral wavefronts [ABSW92].

In our experiment, we use both the polarization and the OAM ($m = \pm 1$) of photons to prepare quantum states of dimension up to 4, spanned by the orthogonal vectors $|H, +1\rangle$, $|H, -1\rangle$, $|V, +1\rangle$ and $|V, -1\rangle$, where $|H, \pm 1\rangle$ ($|V, \pm 1\rangle$) denotes a horizontally (vertically) polarized photon with OAM $m = \pm 1$. We first generate via spontaneous parametric down-conversion (SPDC) pairs of photons (signal and idler) entangled in both polarization and OAM, see Fig. 3.5. The entangled state is of the form $|\Psi^-\rangle_{\text{pol}} \otimes |\Psi^-\rangle_{\text{OAM}}$, where $|\Psi^-\rangle_{\text{pol}} = \frac{1}{\sqrt{2}}(|H\rangle_s|V\rangle_i - |V\rangle_s|H\rangle_i)$ and $|\Psi^-\rangle_{\text{OAM}} = \frac{1}{\sqrt{2}}(|m = 1\rangle_s|m = -1\rangle_i + |m = -1\rangle_s|m = 1\rangle_i)$. By performing a projective measurement on the idler photon, we prepare the signal photon in a well-defined state of polarization and OAM. In particular, we project the idler photon on states of the form $(\cos\theta|H\rangle_i + \sin\theta|V\rangle_i) \otimes |m \pm 1\rangle_i$, which has the effect of preparing the signal photon in the state $(\sin\theta|H\rangle_s - \cos\theta|V\rangle_s) \otimes |m \mp 1\rangle_s$. Thus the combination of the source of entanglement and the measurement of the idler photon represents the state preparator. The prepared state is encoded on the signal photon which is then measured. The signal photon represents the mediating particle between the state preparator and measurement device of Fig. 3.1.

To implement a continuous transition from quantum to classical states, a polarization-dependent temporal delay τ between the signal and idler photons is introduced. If the temporal delay between the photons exceeds their correlation time, the coherence is lost, i.e., the off-diagonal terms vanish for all states in the ensemble (see Supp. Info in [HGM$^+$12]).

For the sake of clarity, we list the assumptions made when processing the observed data: (i) the statistical behavior $P(B|X, Y)$ is the same at every run of the experiment; (ii) the detectors used in the source and in the measurement device do not share prior correlations. This implies, as shown in Thm. 3.7, that one does not need to take in account the no-click events due to imperfect detectors for dimension $d = 2$. For systems of higher dimensionality, we need to assume that the set of observed events is a random sample of the whole set of events obtained with perfect detectors. This is known as the fair sampling assumption. (iii) the observer can freely choose the preparation and measurement in each run. All these assumptions are standard in any estimation scenario. The value of the dimension witnesses is then calculated from the raw data, that is, from all the observed coincidences between detection at the preparator and at the measuring device, including

dark counts.

In the experiment we first generate and measure the four qubit states $|\phi_x\rangle$ given in Fig. 3.5. The first measurement ($y = 1$) assigns dichotomic measurement results of $b = +1$ and $b = -1$ to horizontally and vertically polarized photons, respectively. The second measurement ($y = 2$) assigns $b = +1$ and $b = -1$ to OAM values of $m = +1$ and $m = -1$, respectively. The third measurement ($y = 3$) assigns $b = +1$ and $b = -1$ to photons polarized at $+45^o$ and -45°, respectively. The expected value of the dimension witness of Eq. 3.23 for this combination of states and measurements is $I_4 = 3+2\sqrt{2} \sim 5.83$ (see Supp. Info. in [HGM+12]). From our experimental data we obtain $I_4 = 5.66 \pm 0.15$. This clearly demonstrates the quantum nature of our 2-dimensional system, since classical bits always satisfy $I_4 \leq 5$.

In the above, the delay between signal and idler photons was set to $\tau = 0$. Now we gradually increase this delay to convert a qubit into a classical bit. The measured value of the witness I_4 then drops below 5, as expected (see blue triangles of Fig. 3.6).

Next we generate ensembles of qutrits. The prepared states and the measurements are identical to the previous (qubit) experiment, except that the OAM of state $|\phi_3\rangle$ is now flipped. For $\tau = 0$, we obtain a measured value of the witness of $I_4 = 7.57 \pm 0.13$, certifying the presence of a quantum system of dimension (at least) 3. This value is in good agreement with the theoretical prediction of $I_4 = 5 + 2\sqrt{2} \sim 7.83$ for this set of states and measurements.

Now, increasing again delay τ between the photons, the value of the witness drops below 7. In a certain range of delays, the value of I_4 remains above the qubit bound of 6, testifying that at least 3 dimensions are present (see red circles of Fig. 3.6). In the limit of large delays, the values of the witness are still larger than the bound of $I_4 = 5$ for bits, but below the bound for qubits. This is because the curve was measured with a set a measurements optimized for the qutrit/trit discrimination. This set of measurements is not optimum for the trit/qubit discrimination.

Finally, we prepare classical 4-dimensional systems, i.e. quarts. Now the first measurement ($y = 1$) assigns the outcome $b = -1$ to vertically polarized photons with OAM $m = -1$, and the outcome $b = +1$ to all the other orthogonal states. The second measurement ($y = 2$) assigns $b = +1$ and $b = -1$ to horizontally and vertically polarized photons, respectively. The third measurement ($y = 3$) assigns $b = +1$ and $b = -1$ to OAM of $m = +1$ and $m = -1$, respectively. In this case, the expected value of the witness is $I_4 = 9$, which corresponds to the algebraic maximum. Experimentally we measure $I_4 = 8.57 \pm 0.06$, which violates the qutrit bound

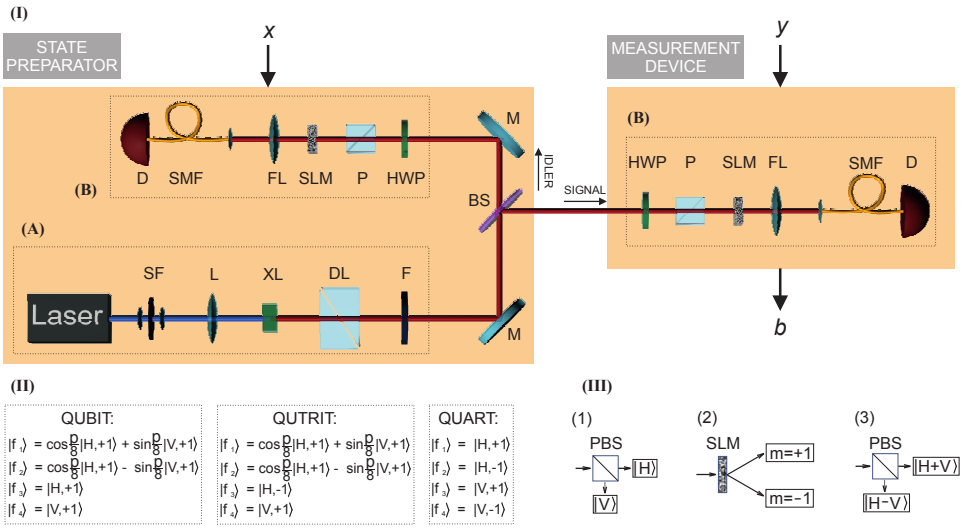

Figure 3.5: Experimental setup. (I) The state preparator consists of a source of entangled photons (A), followed by a measurement (B) on one photon of the pair (idler) that prepares its twin photon (signal) in the desired state. The signal photon is then sent to the measurement device. Block (A) is the source of entangled photons. The second harmonic (Inspire Blue, Spectra Physics/Radiantis) at a wavelength of 405 nm of a Ti:saphire laser in the picosecond regime (Mira, Coherent) is shaped by a spatial filter and focused into a 1.5-mm thick crystal of beta-barium borate (BBO), where SPDC takes place. The nonlinear crystal is cut for collinear type-II down-conversion so that the generated photons have orthogonal polarizations. Before splitting the signal and idler photon, a polarization-dependent temporal delay τ is introduced. The delay line (DL) consists of two quartz prisms whose mutual position determines the difference between the propagation times of photons with orthogonal polarizations. Block (B) performs a measurement on the idler photon to prepare the signal photon. It consists of a half-wave plate (HWP), polarizer (P), spatial-light modulator (SLM) and a Fourier-transform lens (FL). The half-wave plate and polarizer project the photon into the desired polarization state. The desired OAM state is selected by the SLM. SLM encodes computer-generated holograms that transform the $m = +1$ state or $m = -1$ state into the fundamental LG state LG_{00} [MTTT07] that is coupled into a single-mode fiber (SMF). The measurement device uses an identical block (B) to measure the signal photon. (II) Ensembles of quantum states $|\phi_x\rangle$ ($x = 1...4$) prepared in the experiment. (III) Measurements performed at the measurement device for qubits and qutrits. In the case of quarts, the three measurements are constructed by combining (1) and (2).

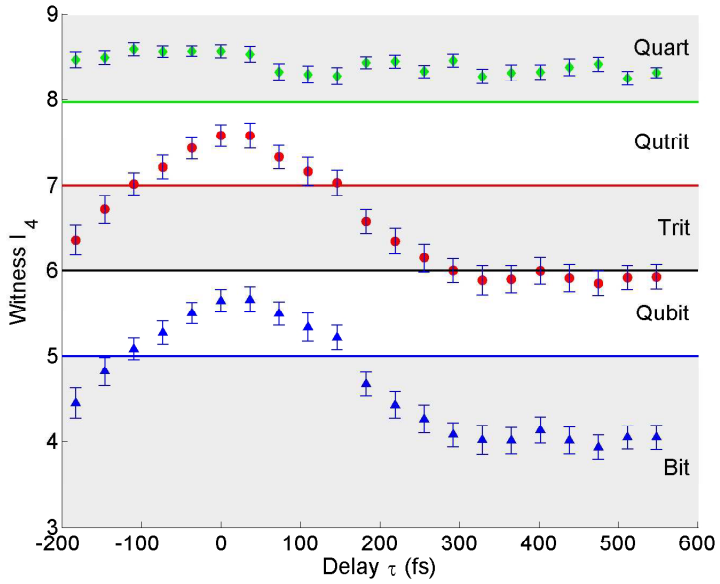

Figure 3.6: Dimension witness I_4 for qubit (blue triangles), qutrit (red circles) and quart (green diamonds) as a function of temporal delay τ. For delays $\tau > 255$ fs, coherence is lost and quantum superpositions turn into statistical mixtures, i.e., classical states. The maximum observed violations for qubit, qutrit and quart are 5.66 ± 0.15, 7.57 ± 0.13, and 8.57 ± 0.06, respectively. These values are close to the corresponding theoretical bounds, given in Table 3.1, which are represented here by the horizontal lines. The error bars plot standard deviations on the value of the witness calculated from the measured data using error propagation rules.

of $I_4 = 7.97$ by more than 10 standard deviations. In this case, the values of the witness are independent of the temporal delay τ. This is because the state is here classical (a statistical mixture of orthogonal quantum states) and no superposition is present (see green diamonds of Fig. 3.6).

To conclude, we have demonstrated how the concept of dimensionality, which is fundamental in science, can be experimentally tested. Using dimension witnesses, we have bounded the dimension of classical and quantum systems only from measurement statistics, without any assumption on the internal working of the devices used in the experiment. Dimension witnesses represent an example of a device-independent estimation technique, in which relevant information about an unknown system is obtained solely from the measurement data. Device-independent techniques provide an alternative approach to existing quantum estimation techniques, such as quantum tomography or entanglement witnesses, which crucially rely on assumptions that may be questionable in complex setups, e.g., its Hilbert space dimension. Our implementation demonstrates how the device-independent approach can be employed to experimentally estimate the dimension of an unknown system.

Chapter 4

Maximally nonlocal quantum correlations

Since the seminal work by Bell [Bel64], we know that there exist quantum correlations that cannot be thought of classically. This impossibility is known as nonlocality and follows from the fact that the correlations obtained when performing local measurements on entangled quantum states may violate a Bell inequality, which sets conditions satisfied by all classically correlated systems.

In this section we study nonlocality from a quantitative point of view. The local content is defined as the fraction of events that admit a description in terms of local hidden variables. We study the local content as a quantifier of nonlocality and find quantum correlations that are maximally nonlocal.

In Section 4.1 we derive a recipe to find maximally nonlocal correlations. We relate these correlations to Kochen-Specker theorems, which study classical assignments to quantum operators at the single-party level. We apply our techniques to design an experiment with highly nonlocal correlations. We report on experimental results, yielding the most nonlocal correlations ever reported.

In Section 4.2 we study maximally nonlocal correlations in the multipartite scenaro. Furthermore, we find correlations that are maximally nonlocal, monogamous and random. These properties make them suitable for multipartite device-independent information protocols. Indeed, we show that the correlations can be applied to protocols of device-independent secret sharing.

4.1 Bounding the local content of quantum correlations

The standard nonlocality scenario consists of two distant systems on which two observers, Alice and Bob, perform respectively M_a and M_b different measurements of d_a and d_b possible outcomes. The outcomes of Alice and Bob are respectively labeled a and b, while their measurement choices are x and y, with $a = \{0, \ldots, d_a - 1\}$, $b = 0, \ldots, d_b - 1$, $x = 1, \ldots, M_a$, and $y = 1, \ldots, M_b$. The correlations between the two systems are encapsulated in the joint conditional probability distribution $P(A, B|X, Y)$, a vector with entries $P(a, b|x, y)$ being the probability of obtaining outcomes a and b when measurements x and y have been performed by Alice and Bob respectively.

As explained in sections 2.1.2 and 2.2.4, the violation of Bell inequalities by entangled states implies that $\mathcal{L} \subset \mathcal{Q}$. A similar gap, $\mathcal{Q} \subset \mathcal{P}$, appears when considering quantum versus general nonsignaling correlations: there exist correlations that, despite being compatible with the no-signaling principle, cannot be obtained by performing local measurements on any quantum system [PR94]. In particular, there exist nonsignaling correlations that exhibit stronger nonlocality, in the sense of giving larger Bell violations, than any quantum correlations.

Interestingly, there are situations in which this second gap disappears: quantum correlations are then maximally nonlocal, as they are able to attain the maximal Bell violation compatible with the no-signaling principle. Geometrically, in these extremal situations quantum correlations reach the border of the set of nonsignaling correlations. From a quantitative point of view, it is possible to detect this effect by computing the local content [EPR92] of the correlations, see section 2.2.2. This quantity measures the fraction of events that can be described by a local model. Maximally nonlocal correlations feature $p_L = 0$ [see Fig. 2.4].

Any Bell violation provides an upper bound on the local fraction of the correlations that cause it. In fact, a Bell inequality is defined as $\beta \cdot P(A, B|X, Y) = \sum \beta_{a,b,x,y} P(a, b|x, y) \leq \beta_L$, where $\beta_{a,b,x,y}$ is a tensor of real coefficients. The maximal value of the left-hand side of this inequality over classical correlations defines the local bound β_L, whereas its maximum over quantum and nonsignaling correlations gives the maximal quantum and nonsignaling values β_Q and β_{NS}, respectively. From this and Def. 2.3 it follows immediately that [BKP06]

$$p_L \leq \frac{\beta_{NS} - \beta_Q}{\beta_{NS} - \beta_L} = p_{L max}. \tag{4.1}$$

Thus, quantum correlations violating a Bell inequality as much as any nonsignaling correlations feature $p_L = 0$.

Here we study the link between the Kochen-Specker (KS) [KS67] and Bell's theorems, previously considered in Refs. [HR83, Cab01, BBT05, CK06, HHH$^+$10]. We recast this link in the form of Bell inequalities maximally violated by quantum states. We then show that the resulting Bell inequalities can be used to get experimental bounds on the nonlocal content of quantum correlations that are significantly better than Bell tests based on more standard Bell inequalities or multipartite Greenberger-Horne-Zeilinger (GHZ) paradoxes [GHZ89]. This allows us to perform an experimental demonstration, which yields an experimental upper bound on the local part $p_{L_{max}} = 0.218 \pm 0.014$. To our knowledge, this represents the lowest value ever reported, even taking into account multipartite Bell tests.

4.1.1 General formalism

In this section, we present the details of the construction to derive different Bell inequalities maximally violated by quantum mechanics from every proof of the KS theorem. This construction was first introduced in [HR83] and was later applied in the context of "all-versus-nothing" nonlocality tests [Cab01], pseudo-telepathy games (see [BBT05] and references therein), the free will theorem [CK06], and quantum key distribution [HHH$^+$10]. Here we exploit it to generate quantum correlations with no local part.

Kochen-Specker theorem

The Kochen-Specker (KS) theorem studies whether deterministic outcomes can be assigned to von Neumann quantum measurements, in contrast to the quantum formalism which can only assign probabilities.

Definition 4.1. ***The Kochen-Specker set.*** *A set of p rank-1 projectors $\{\tilde{\Pi}_1, \ldots, \tilde{\Pi}_p\}$ is a KS set if there exist no map $v(\tilde{\Pi}_j) = \{0, 1\}$ with the property that $\sum_{\tilde{\Pi}_j \in B_z} v(\tilde{\Pi}_j) = 1$ for every subset of projectors B_z being an orthonormal basis.*

Let us reformulate it to give a more precise interpretation to the KS set. Let us denote by M the number of orthonormal basis that one can construct from the set $\{\tilde{\Pi}_1, \ldots, \tilde{\Pi}_p\}$. Each of these basis defines a von Neumann measurement z defined by a set of d orthogonal projectors acting on a Hilbert space of dimension d. Therefore, one can construct M such measurements

from the KS set. Let us rename the projectors involved in the M measurements as Π_i^z, with $z = 1, \ldots, M$ and $i = 1, \ldots, d$, such that $\Pi_i^z \Pi_{i'}^z = \delta_{i,i'}$ and $\sum_i \Pi_i^z = \mathbb{I}$ for all z, with $\mathbb{I}$ being the identity operator. The measurements may have some common projectors, so let us denote by D_j the set of two-tuples $D_j = \{(i, z)\}$ such that $(i, z) \in D_j$ if $\Pi_i^z = \tilde{\Pi}_j$. Assigning deterministic values to the M von Neuman measurements is equivalent to finding a map $f(z) \in \{1, \ldots, d\}$, or equivalently a map $f(z, i) \in \{0, 1\}$ such that $\sum_i f(z, i) = 1$. If one imposes no restriction, such assignment is always possible. However, it is not the case if one imposes an extra condition commonly referred to as noncontextuality: the assignment f has to be such that it assigns equal values to outcomes represented by equal projectors. That is, $f(z, i) = f(z', i')$ if $\Pi_i^z = \Pi_{i'}^{z'}$. Now the KS set plays a significant role: if $\{\tilde{\Pi}_1, \ldots, \tilde{\Pi}_p\}$ is a KS set, then this is impossible to find such map f, otherwise there would exist a map v simply defined as $v(\tilde{\Pi}_j) = f(z, i)$.

One can find in the literature several examples of KS sets [KS67, Mer90b, Per91, CEGA96] which show that a noncontextual assignment to quantum measurements is impossible. Therefore quantum mechanics is said to be a contextual theory.

Each KS set leads to maximally nonlocal correlations

Let us now see how this highly nontrivial configuration of measurements given by a KS set can be used to derive maximally nonlocal quantum correlations. Consider the standard Bell scenario depicted in Fig. 4.1 (b). Two distant observers (Alice and Bob) perform uncharacterized measurements in a device-independent scenario. Let us assume that Alice can choose among $M_a = M$ measurements of $d_a = d$ outcomes. On the other hand, Bob can choose among $M_b = p$ measurements of $d_b = 2$ outcomes, labeled by 0 and 1. We denote Alice's (Bob's) measurement choice by x (y) and her (his) outcome by a (b) and the joint probability distribution by $P(A, B|X, Y)$.

Theorem 4.2. *From every KS set $\{\tilde{\Pi}_1, \ldots, \tilde{\Pi}_p\}$ one can derive a nonlocal experiment yielding a quantum probability distribution $P(A, B|X, Y)$ with zero local content.*

Proof. The strategy of the proof can be summarized as follows. We will show that there exist a Bell inequality $\beta \cdot P(A, B|X, Y)$ with two properties: (i) there exist a quantum realization of the experiment yielding correlations $P_Q(A, B|X, Y)$, such that $\beta \cdot P_Q(A, B|X, Y) \equiv \beta_Q = \beta_{NS}$ and (ii) the maximum over local probability distributions fulfills $\beta_L = \beta_{NS} - 1$. Therefore, using (4.1) one gets that $p_L = 0$.

Let us show (i): Consider the following quantum realization of the experiment yielding $P_Q(A,B|X,Y)$: Alice and Bob perform their measurements on the bipartite maximally entangled state $|\psi_d\rangle = \sum_{k=0}^{d-1} \frac{1}{\sqrt{d}}|kk\rangle$. Alice performs the M measurements that correspond with the M different orthonormal basis that one can construct out of the KS set. The measurements are labelled such that when Alice chooses input x, measurement $\{\Pi_a^x\}_{a=[1:d]}$ is performed. In turn, when Bob chooses input y, the following 2-outcome measurement takes place: $\{(\tilde{\Pi}_y)^*, \mathbb{I}-(\tilde{\Pi}_y)^*\}$, where the asterisk (*) denotes complex conjugation, and where the first and second projectors are assigned to outcomes 0 and 1 respectively. This realization leads to the nonsignaling value β_{NS} of the following linear combination of probabilities:

$$\begin{aligned}\beta \cdot P(A,B|X,Y) &= \sum_{y=1}^{p} \sum_{(a',x)\in D_y} [P(a=a', b=1|x,y) \\ &+ P(a\neq a', b=0|x,y)]. \end{aligned} \tag{4.2}$$

Indeed, for all the terms appearing in (4.2), $P_Q(a=a', b=1|x,y) + P_Q(a\neq a', b=0|x,y) = 1$. This can be easily seen by noticing that if Bob's output is equal to 1, Alice's system is projected onto $\tilde{\Pi}_y = \Pi_{a'}^x$, and thus, the result of Alice's measurement x is a'. On the contrary, if Bob's box outputs 0, Alice's system is projected onto $\mathbb{I}-\tilde{\Pi}_y = \mathbb{I}-\Pi_{a'}^x$, and thus, Alice's outcome is such that $a\neq a'$. As the sum of the two probabilities $P_Q(a=a', b=1|x,y)$ and $P_Q(a\neq a', b=0|x,y)$ can never be larger than 1, one has

$$\beta_Q = \beta_{NS} = \sum_{y=1}^{p} \sum_{(a',x)\in D_y} 1. \tag{4.3}$$

Let us show (ii): As for local correlations, we now show that it is $\beta_L \leq \beta_{NS} - 1$. To see this, recall first that the maximum of (4.2) over local models is always reached by some deterministic model, in which a deterministic outcome is assigned to every measurement [and all probabilities in (4.2) can thus only be equal to 0 or 1]. Hence, deterministic models can only feature $\beta_L \in \mathbb{Z}$. Therefore, it suffices to show that the maximum of (4.2) over local models satisfies $\beta_L < \beta_{NS}$. This can be proven by *reductio ad absurdum*. Suppose that a local deterministic model attains the value β_{NS}. Any deterministic model then specifies the outcomes a and b on both sides for all measurements. Equivalently, it can be understood as a definite assignment to every measurement outcome on Alice's and Bob's sides: $v_A(\Pi_a^x) \in \{1,0\}$ and $v_B(\tilde{\Pi}_y) \in \{1,0\}$, with $\sum_a v_A(\Pi_a^x) = 1$, for all x. If

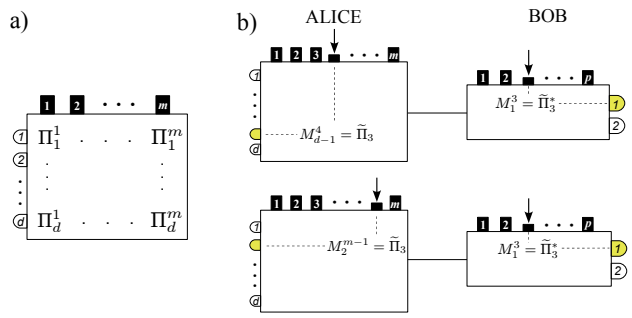

Figure 4.1: **Noncontextual assignments in the black-box scenario.** (a) A KS proof consists of a single observer, say Alice, who performs m measurements of d outcomes. The KS proof requires that outcomes of different measurements correspond to the same projector. There are altogether p projectors, denoted by $\tilde{\Pi}_j$, shared by different measurements. The common projectors impose constraints that, if the outcomes are assigned by noncontextual deterministic maps, lead to contradictions. (b) In the Bell test associated with the KS proof, Bob's box has $m_b = p$ possible measurements of $d_b = 2$ outcomes. In the quantum setting, the two observers share a maximally entangled state. Alice makes $m_a = m$ measurements of $d_a = d$ outcomes, which correspond to the observables in the KS proof. Bob's measurements are perfectly correlated with the p projectors $\tilde{\Pi}_j$ on Alice's side, thanks to the properties of the maximally entangled state. A local model reproducing all these correlations would imply the existence of a deterministic assignment for Bob's measurements, which is impossible as they define a KS set.

(4.2) reaches its maximum algebraic value, the assignment map is subject to the constraints $v_A(\Pi_a^x) = v_B(\tilde{\Pi}_y) = v_A(M_{a'}^{x'})$ for all (a, x) and $(a', x') \in D_y$. This implies that $\sum_{\tilde{\Pi}_j \in B_y} v_B(\tilde{\Pi}_j) = 1$ for all the M orthonormal basis. This, however, is prohibited because $\{\tilde{\Pi}_y\}$ is a KS set. Thus, one concludes that $\beta_L \leq \beta_{NS} - 1$. This completes the proof. $\square$

Before concluding this section, we would like to emphasize that this recipe can lead to other, possibly nonequivalent, Bell inequalities. For instance, it is possible to keep Alice's measurements equal to those in the KS proof and replicate them on Bob's side, *i.e.*, $\{\Pi_b^y = (\Pi_a^x)^*$, with $y = x$ and $b = a\}$. Note that then all the projectors needed to enforce the KS constraints on Alice's side by means of perfect correlations appear on Bob's side. Other examples are provided by some proofs that possess inherent symmetries, allowing for peculiar distributions of the contexts in the proof between Alice's and Bob's sides, as is discussed in the next section.

	$y=1$	$y=2$	$y=3$	
$x=1$	Z_2	X_1	X_1Z_2	$=\mathbb{I}$
$x=2$	Z_1	X_2	Z_1X_2	$=\mathbb{I}$
$x=3$	Z_1Z_2	X_1X_2	Y_1Y_2	$=-\mathbb{I}$
	$=\mathbb{I}$	$=\mathbb{I}$	$=\mathbb{I}$	$\prod$

Table 4.1: **The Peres-Mermin square**. One of the simplest KS proofs was derived by Peres and Mermin [Per91, Mer90b] and is based on the nine observables of this table. The observables are grouped into six groups of three, arranged along columns and rows. X_n, Y_n, and Z_n refer to Pauli matrices acting on qubits $n=1$ and $n=2$, which span a four-dimensional Hilbert space. Each group constitutes a complete set of mutually commuting (and therefore compatible) observables, defining thus a context. In this way, there are six contexts, and every observable belongs to two different ones. The product of all three observables in each context is equal to the identity $\mathbb{I}$, except for those of the third row, whose product gives $-\mathbb{I}$. It is impossible to assign numerical values 1 or -1 to each one of these nine observables in a way that the values obey the same multiplication rules as the observables. This, in turn, implies that it is impossible to make a noncontextual assignment to the 24 underlying projectors (not shown) in the table (one common eigenbasis per context, with four eigenvectors each).

4.1.2 A simple Bell inequality

The previous recipe is fully general. In this section, in contrast, we apply the ideas just presented to derive a specific Bell inequality maximally violated by quantum mechanics from one of the most elegant KS proofs, introduced by Peres and Mermin [Mer90b, Per91]. Apart from being one of the simplest Bell inequalities having this property, its derivation shows how symmetries in the KS proof can be exploited to simplify the previous construction.

The Peres-Mermin (PM) KS proof is based on the set of observables of Table 4.1, also known as the PM square, which can take two possible values, ± 1. This proof in terms of observables can be mapped into a proof in terms of 24 rank-1 projectors [Per91, CEGA96]. To these projectors we could then apply the formalism of the previous section and derive Bell inequalities maximally violated by quantum correlations of the sort of (4.2). However, some special features of this particular KS proof allow one to simplify the process and derive a simpler inequality straight from the observables. The key point is that in the PM square each operator appears in two different contexts, one being a row and the other a column. This allows one to distribute the contexts between Alice and Bob in such a way that Alice (Bob) performs the measurements corresponding to the rows (columns) (see also [HHH$^+$10]). The corresponding Bell scenario, then, is such that Alice and

Bob can choose among three different measurements $x, y \in \{1, 2, 3\}$ of four different outcomes, $a, b \in \{1, 2, 3, 4\}$. Consistent with the PM square, we associate in what follows Alice and Bob's observables x and y with the rows and columns of the square, respectively, and divide the four-value outputs into two bits, $a = (a_1, a_2)$ and $b = (b_1, b_2)$, each of which can take the values ± 1.

Consider first the following quantum realization: Alice and Bob share two two-qubit maximally entangled states $|\psi_4\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{12} \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{34}$, which is equivalent to a maximally entangled state of two four-dimensional systems. Alice possesses systems 1 and 3, and Bob possesses systems 2 and 4. Alice can choose among three different measurements that correspond to the three rows appearing in Table 4.1. If Alice chooses input x, the quantum measurement defined by observables placed in row x is performed. Note that the measurement acts on a four-dimensional quantum state; thus there exist four possible outcomes (one for each eigenvector common to all three observables), which in our scenario are decomposed into two dichotomic outputs. We define a_i to be the value of the observable placed in column $y = i$ for $i = 1, 2$. The value of the third observable in the same row is redundant as it can be obtained as a function of the other two. Equivalently, Bob can choose among three measurements that correspond to the three columns appearing in Table 4.1. If Bob chooses input y, outputs b_j are the values of observables placed in column y and row $x = j$ for $y = 1, 2, 3$ and $j = 1, 2$. This realization attains the algebraic maximum $\beta_Q = \beta_{NS} = 9$ of the linear combination

$$\begin{aligned} \beta &= \langle a_1 b_1 | 1, 1\rangle + \langle a_2 b_1 | 1, 2\rangle + \langle a_1 b_2 | 2, 1\rangle \\ &+ \langle a_2 b_2 | 2, 2\rangle + \langle a_1 a_2 b_1 | 1, 3\rangle + \langle a_1 a_2 b_2 | 2, 3\rangle \\ &+ \langle a_1 b_1 b_2 | 3, 1\rangle + \langle a_2 b_1 b_2 | 3, 2\rangle \\ &- \langle a_1 a_2 b_1 b_2 | 3, 3\rangle, \end{aligned} \tag{4.4}$$

where $\langle f(a_1, a_2, b_1, b_2) | x, y\rangle$ denotes the expectation value of a function f of the output bits for the measurements x and y.

To prove this statement, let us first focus on the term $\langle a_1 b_1 | 1, 1\rangle$. Bit b_1 is obtained as the outcome of the measurement of the quantum observable $Z_4 \otimes \mathbb{I}_2$. As the measurement is performed on the maximally entangled state, the state on Alice's side is effectively projected after Bob's measurement onto the eigenspace of $Z_3 \otimes \mathbb{I}_1$ with eigenvalue b_1. Bit a_1 is defined precisely as the outcome of the measurement of the observable $Z_3 \otimes \mathbb{I}_1$; thus $a_1 = b_1$ and $\langle a_1 b_1 | 1, 1\rangle = 1$. The same argument applies to the first four terms in (4.4). Consider now the term $\langle a_1 a_2 \cdot b_1 | 1, 3\rangle$. Bit b_1 is the outcome of the

measurement of the observable $Z_4 \otimes X_2$. The state after Bob's measurement is effectively projected on Alice's side onto the eigenspace of $Z_3 \otimes X_1$ with eigenvalue b_1. Bit $a_1 \cdot a_2$ is obtained as the measurement output of the observable $Z_3 \otimes X_1$; thus $a_1 \cdot a_2 = b_1$ and $\langle a_1 a_2 \cdot b_1 | 1, 3\rangle = 1$. The same argument applies to the four terms involving products of three bits. The last term $\langle a_1 a_2 \cdot b_1 \cdot b_2 | 3, 3\rangle$ requires a similar argument. Bit $a_1 \cdot a_2$ is obtained as the output of the operator $Y_3 \otimes Y_1$ (note that the product of the observables associated with a_1 and a_2 is $Y_3 \otimes Y_1$, see Table 4.1). Thus the state is effectively projected onto the eigenspace of $Y_4 \otimes Y_2$ with eigenvalue $a_1 \cdot a_2$. The bit $b_1 \cdot b_2$ is precisely the meqasurement outcome of $-Y_4 \otimes Y_2$, thus $a_1 \cdot a_2 = -b_1 \cdot b_2$ and $\langle a_1 a_2 \cdot b_1 \cdot b_2 | 3, 3\rangle = -1$.

We move next to the classical domain, to show that the maximum value of polynomial (4.4) attainable by any local model is $\beta_L = 7$, and thus, the inequality

$$\beta \leq 7, \tag{4.5}$$

with β defined by (4.4), constitutes a valid Bell inequality, maximally violated by quantum mechanics. Remarkably, this inequality has already appeared in Ref. [Cab01] in the context of all-versus-nothing nonlocality tests. Computing the local bound $\beta_L = 7$ can easily be performed by brute force (that is, by explicitly calculating the value of β_L for all possible assignments). However, it is also possible to derive it using arguments similar to those in the previous section. In the PM square, each of the nine dichotomic observables belongs to two different contexts, one being a row and the other a column, as mentioned. Therefore, nine correlation terms are needed to enforce the KS constraints. As said, the symmetries of the PM square allow one to split the contexts between Alice and Bob, arranging these correlation terms in a distributed manner. Such correlation terms correspond precisely to the nine terms appearing in (4.4). Again, the existence of a local model saturating all these terms would imply the existence of a noncontextual model for the PM square, which is impossible.

4.1.3 Previous bounds on local content using other Bell inequalities

The scope of this section is to show how the previous construction offers important experimental advantages when deriving bounds on the local content of quantum correlations. First of all, and contrary to some of the examples of quantum correlations with no local part [BKP06], the Bell inequalities derived here not only involve a finite number of measurements but are in addition resistant to noise. Moreover, as shown in what follows, they allow

one to obtain experimental bounds on the nonlocal part that are significantly better than those based on other Bell tests.

Let us first consider the Collins-Gisin-Linden-Massar-Popescu inequalities presented in [CGL$^+$02]. These inequalities are defined for two measurements of d outcomes. The maximal nonsignaling violation of these inequalities is equal to β_{NS} =4, while the local bound is $\beta_L = 2$. The maximal quantum violation of these inequalities is only known for small values of d [ADGL02, NPA08]. A numerical guess for the maximal quantum violation for any d was provided in [ZG08]. This guess reproduces the known values for small d and tends to the nonsignaling value when $d \to \infty$. Assuming the validity of this guess, a bound on the local content comparable to the experimental value reported in the next section, namely, $p_{L_{max}} = 0.218 \pm 0.014$, requires a number of outputs of the order of 200 (see [ZG08]), even in the ideal noise-free situation. Note that the known quantum realization attaining this value involves systems of dimension equal to the number of outputs, that is, 200, and the form of the quantum state is rather complicated. If the quantum state is imposed to be maximally entangled, the maximal quantum violation tends to 2.9681, which provides a bound on the local content of just $p_{L_{max}} \approx 0.5195$.

The chained inequalities [BC90, BKP06], defined in a scenario where Alice and Bob can both perform m measurements of d outcomes, provide a bound on the local content that tends to zero with the number of measurements, $m \to \infty$ [BKP06]. However, in this limit the nonlocality of the corresponding quantum correlations is not resistant to noise (see Fig. 4.2), and thus, the use of many measurements requires an almost-noise-free realization. We compare the chained inequalities [BC90] for $d = 2$ (the simplest case to implement) with our inequality (4.5) in a realistic noisy situation. The quantum state is written as the mixture of the maximally entangled state, as this state provides the maximal quantum violation of both the chained inequality and inequality (4.5), with white noise,

$$\rho = V|\psi_d\rangle\langle\psi_d| + (1-V)\frac{\mathbb{I}}{d^2}. \tag{4.6}$$

The amount of white noise on the state is quantified by $1-V$. The bound on the local content then reads $p_{L_{max}} = \frac{\beta_{NS} - V\beta_Q + (1-V)\beta_{\mathbb{I}}}{\beta_{NS} - \beta_L}$, where $\beta_{\mathbb{I}}$ is the value of the Bell inequality given by white noise with the optimal measurements. We plot the obtained results in Fig. 4.2. As shown there, the Bell inequality considered here provides better bounds on the local content than the chained inequalities for almost any value of the noise.

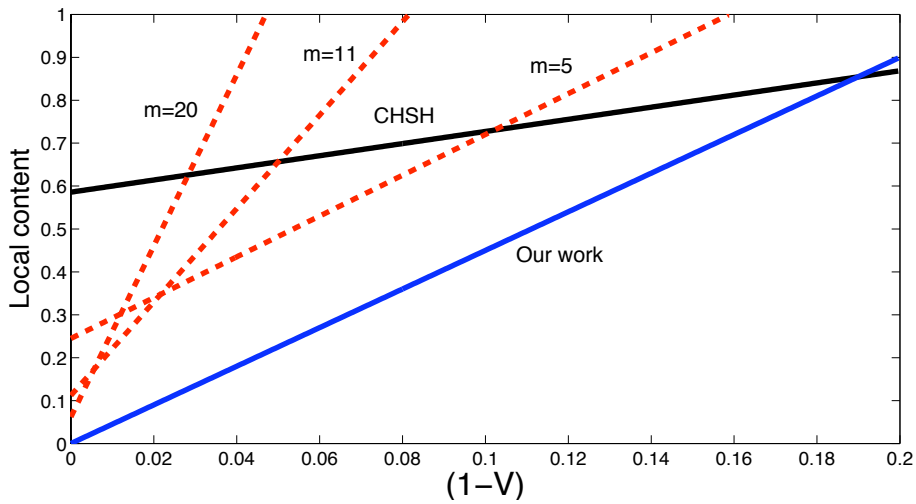

Figure 4.2: **Resistance to noise of different Bell tests.** Dashed red curves show the resistance to noise of the chained inequality [BC90] for different numbers m of measurements. The local content and also the resistance to noise tend to zero when the number of measurements tends to infinity, as expected. Standard Bell inequalities, such as the Clauser-Horne-Shimony-Holt (CHSH) inequality [CHSH69], can be violated in a robust manner and with few measurements, but the obtained bound on the local content never goes to zero (in fact, the CHSH inequality is the chained inequality [BC90] for $m = 2$). Inequality (4.5) (solid blue curve) in contrast combines all three features: its violation is resistant to noise and requires few measurements, and its bound on the local content is equal to zero in the noise-free case.

4.1.4 Experimental highly nonlocal quantum correlations

We performed a test of inequality (4.5) with two entangled photons, A and B, generated by spontaneous parametric down conversion (SPDC). We used type-I phase matching with a β-barium-borate (BBO) crystal. The source used a single crystal and a double passage of the UV beam after the reflection on a spherical mirror [see Fig. 4.3 (a)] and generated the hyperentangled state [CBP+05]

$$\begin{aligned} |\Psi\rangle &= \frac{1}{\sqrt{2}}(|H\rangle_A|H\rangle_B + |V\rangle_A|V\rangle_B) \\ &\quad \otimes \frac{1}{\sqrt{2}}(|r\rangle_A|l\rangle_B + |l\rangle_A|r\rangle_B)\,, \end{aligned} \tag{4.7}$$

where $|H\rangle$ ($|V\rangle$) represents the horizontal (vertical) polarization and $|r\rangle$ and $|l\rangle$ are the two spatial path modes in which each photon can be emitted.

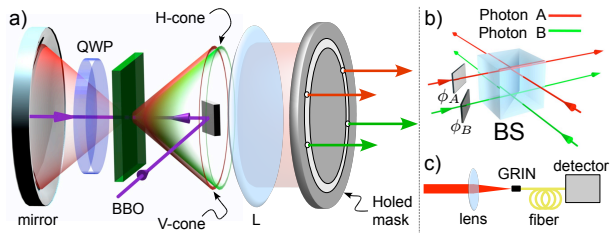

Figure 4.3: **Experimental setup.** (a) Source of hyperentangled photon states. The relative phase between the states $|HH\rangle_{AB}$ and $|VV\rangle_{AB}$ can be varied by translating the spherical mirror. A lens L located at a focal distance from the crystal transforms the conical emission into a cylindrical one. (b) Scheme for the path measurements. **c.** The parametric radiation is coupled into single-mode fibers by a GRIN lens and sent to the detectors.

Maximally entangled state $|\psi_4\rangle$ between A and B, as defined in Sec. 4.1.2, is recovered from (4.7) through the following identification: $|H\rangle_{A,B} \equiv |0\rangle_{1,2}$, $|V\rangle_{A,B} \equiv |1\rangle_{1,2}$, $|r\rangle_A \equiv |0\rangle_3$, $|l\rangle_A \equiv |1\rangle_3$, $|l\rangle_B \equiv |0\rangle_4$, and $|r\rangle_B \equiv |1\rangle_4$. Therefore, state (4.7) also allows for the maximal violation of (4.5).

In the SPDC source, the BBO crystal is shined on by a vertically polarized continuous wave (cw) Ar^+ laser ($\lambda_p = 364$ nm), and the two photons are emitted at degenerate wavelength $\lambda = 728$ nm and with horizontal polarization. Polarization entanglement is generated by the double passage (back and forth, after the reflection on a spherical mirror) of the UV beam. The backward emission generates the so called V cone: the SPDC horizontally polarized photons passing twice through the quarter-wave plate (QWP) are transformed into vertically polarized photons. The forward emission generates the H cone [the QWP behaves almost as a half-wave plate (HWP) for the UV beam]. See Fig. 4.3 (a). Thanks to temporal and spatial superposition, the indistinguishability of the two perpendicularly polarized SPDC cones creates polarization entanglement $(|H\rangle_A|H\rangle_B + |V\rangle_A|V\rangle_B)/\sqrt{2}$. The two polarization entangled photons are emitted over symmetrical directions belonging to the surface of the cone. By selecting two pairs of correlated modes by a four-holed mask [CBP+05, VPDMM08, CVB+10] it is possible to generate path entanglement.

In order to measure the path operators, the four modes of the hyperentangled state are matched on a beam splitter (BS) in a complete indistinguishability condition. This operation corresponds to the projection onto $\frac{1}{\sqrt{2}}(|r\rangle_A + e^{i\phi_A}|l\rangle_A) \otimes \frac{1}{\sqrt{2}}(|r\rangle_B + e^{i\phi_B}|l\rangle_B)$. Suitable tilting of two thin glass plates allows one to set phases ϕ_A and ϕ_B [see Fig. 4.3

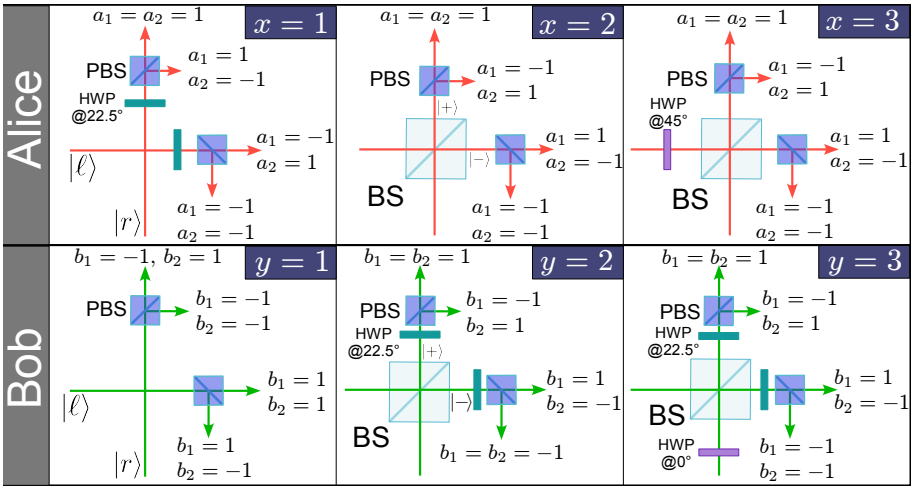

Figure 4.4: **Measurement setups used by Alice and Bob.** See text for a detailed explanation of the measurements. BS, beam splitter; PBS, polarizing beam splitter; HWP, half-wave plate.

(b)]. Photon collection is performed by integrated systems of graded-index lenses and single-mode fibers connected to single-photon counting modules [RVC+09, VDDMM09] [see Fig. 4.3 c)]. Polarization analysis is performed in each output mode by a polarizing beam splitter (PBS) and a properly oriented HWP. The experimental setup used for each polarization measurement setting is shown in Fig. 4.4.

The nine terms of Bell polynomial (4.4) correspond to the different combinations between one of Alice's three contexts and one of Bob's three contexts listed in Table 4.2. In the settings $x = 1, 2$ ($y = 1, 2$) Alice (Bob) must project into states that are separable between path and polarization (eigenstates of Pauli operators X and Z). To project into $\{|r\rangle, |l\rangle\}$ the modes are detected without BS. On the other hand, the BS is used to project into $\frac{1}{\sqrt{2}}(|r\rangle \pm |l\rangle)$. PBSs and wave plates have been exploited to project into $\{|H\rangle, |V\rangle\}$ or $\frac{1}{\sqrt{2}}(|H\rangle \pm |V\rangle)$. More details are needed for contexts $x, y = 3$, corresponding to the projection into single-photon Bell states (the two entangled qubits of the Bell state are encoded in polarization and path of the single particle, see Table 4.2). For instance, let us consider the projection on the states $|H\rangle|l\rangle \pm |V\rangle|r\rangle$ and $|V\rangle|l\rangle \pm |H\rangle|r\rangle$ for Alice. By inserting a HWP oriented at 45° on the mode $|l\rangle_A$ before the BS, the previous states become $|V\rangle|\pm\rangle$ and $|H\rangle|\pm\rangle$, respectively. The two BS outputs allow one to

Alice

	a_1=−1, a_2=−1	a_1=−1, a_2=1	a_1=1, a_2=−1	a_1=1, a_2=1
$x = 1$	$\|-\rangle\|l\rangle$	$\|+\rangle\|l\rangle$	$\|-\rangle\|r\rangle$	$\|+\rangle\|r\rangle$
$x = 2$	$\|V\rangle\|-\rangle$	$\|V\rangle\|+\rangle$	$\|H\rangle\|-\rangle$	$\|H\rangle\|+\rangle$
$x = 3$	$\|H\rangle\|l\rangle - \|V\rangle\|r\rangle$	$\|H\rangle\|l\rangle + \|V\rangle\|r\rangle$	$\|H\rangle\|r\rangle - \|V\rangle\|l\rangle$	$\|H\rangle\|r\rangle + \|V\rangle\|l\rangle$

Bob

	b_1=−1, b_2=−1	b_1=−1, b_2=1	b_1=1, b_2=−1	b_1=1, b_2=1
$y = 1$	$\|V\rangle\|r\rangle$	$\|H\rangle\|r\rangle$	$\|V\rangle\|l\rangle$	$\|H\rangle\|l\rangle$
$y = 2$	$\|-\rangle\|-\rangle$	$\|-\rangle\|+\rangle$	$\|+\rangle\|-\rangle$	$\|+\rangle\|+\rangle$
$y = 3$	$\|+\rangle\|r\rangle - \|-\rangle\|l\rangle$	$\|+\rangle\|r\rangle + \|-\rangle\|l\rangle$	$\|-\rangle\|r\rangle - \|+\rangle\|l\rangle$	$\|-\rangle\|r\rangle + \|+\rangle\|l\rangle$

Table 4.2: **Measurement settings**. Each row represents a measurement (context). The four states in each row represent the four projectors of each measurement. $a_{1,2}$ and $b_{1,2}$ are the two-bit outcomes of Alice and Bob respectively. In each state, the first ket refers to polarization, while the second one refers to path. $|\pm\rangle$ correspond to $\frac{1}{\sqrt{2}}(|H\rangle \pm |V\rangle)$ or $\frac{1}{\sqrt{2}}(|r\rangle \pm |l\rangle)$, for polarization or path respectively.

Correlation	Experimental result
$\langle a_1 b_1 \|1,1\rangle$	0.9968 ± 0.0032
$\langle a_1 b_2 \|2,1\rangle$	0.9759 ± 0.0058
$\langle a_2 b_1 \|1,2\rangle$	0.9645 ± 0.0068
$\langle a_2 b_2 \|2,2\rangle$	0.941 ± 0.010
$\langle a_1 a_2 b_1 \|1,3\rangle$	0.9705 ± 0.0048
$\langle a_1 a_2 b_2 \|2,3\rangle$	0.9702 ± 0.0049
$\langle a_1 b_1 b_2 \|3,1\rangle$	0.9688 ± 0.0073
$\langle a_2 b_1 b_2 \|3,2\rangle$	0.890 ± 0.013
$\langle a_1 a_2 b_1 b_2 \|3,3\rangle$	-0.888 ± 0.018

Table 4.3: **Experimental results**. Errors were calculated by propagating Poissonian errors of the counts.

discriminate between $|r\rangle + |l\rangle$ and $|r\rangle - |l\rangle$, while the two outputs of the PBSs discriminate $|H\rangle$ and $|V\rangle$.

Table 4.3 provides the experimental values of all nine correlations in Bell polynomial (4.4). The obtained violation for Bell inequality (4.5) is

Experiment	p_L
Aspect *et al.* [ADR82]	$\lesssim$ 0.80
Weihs *et al.* [WJS+98]	$\lesssim$ 0.64
Kiesel *et al.* [KSW+05]	$\lesssim$ 0.64
Zhao *et al.* [ZYC+03]	$\lesssim$ 0.60
Pomarico *et al.* [PBS+11b]	$\lesssim$ 0.49
Our experiment (and Yang *et al.* [YZZ+05])	$\lesssim$ 0.22

Table 4.4: **Bounds on the local content of quantum correlations from previous Bell experiments**. The selection includes representative experiments testing different forms of nonlocality, or Bell inequalities, in both the bipartite [CHSH69, BC90] and multipartite [Ard92, SASA05] scenarios. Other published experiments, not shown in the table, lead to $p_{L_{max}} > 0.49$. Note the significant improvement given by the techniques discussed here (see also Sec. 4.1.3).

$\beta_Q^{exp} = 8.564 \pm 0.028$ and provides the upper bound $p_{L_{max}} = 0.218 \pm 0.014$. At this point it is important to mention that another experimental test of (4.5) was reported in Ref. [YZZ+05] in the framework of all-versus-nothing nonlocality tests. The violation in Ref. [YZZ+05] is compatible (within experimental errors) with the value obtained by our experiment.

4.1.5 Conclusions and discussions

In this section we have provided a systematic recipe for obtaining bipartite Bell inequalities from every proof of the Kochen-Specker theorem. These inequalities are violated by quantum correlations in an extremal way, thus revealing the fully nonlocal nature of quantum mechanics. We have shown that these inequalities allow establishing experimental bounds on the local content of quantum correlations that are significantly better than those obtained using other constructions. This enabled us to experimentally demonstrate a Bell violation leading to the highly nonlocal bound $p_L \lesssim 0.22$.

The local content p_L of some correlations $P(A, B|X, Y)$ can be understood as a measure of their locality, as it measures the fraction of experimental runs admitting a local-hidden-variable description. As mentioned, some of the previously known examples of bipartite inequalities featuring fully nonlocal correlations, i.e., $p_L = 0$, for arbitrary dimensions require an infinite number of measurement settings and are not robust against noise [EPR92, BKP06]. More standard Bell inequalities using a finite number of measurements, such as the well-known Clauser-Horne-Shimony-Holt inequality [CHSH69], give a local weight significantly larger than zero even in

the noise-free situation. Thus, the corresponding experimental violations, inevitably noisy, have only managed to provide bounds on the local content not smaller than 0.5 (see Table 4.4). In contrast, the theoretical techniques provided here enable the experimental demonstration of highly nonlocal correlations. This explains why the experimental bound we obtain is significantly better than those of previous Bell tests, even including multipartite ones. In fact, multipartite Greenberger-Horne-Zeilinger tests [GHZ89] also in principle yield $p_L = 0$ [BKP06] using a finite number of measurements and featuring robustness against noise. Still, to our knowledge, the reported experimental violations lead to significantly worse bounds on p_L (see Table 4.4). Our analysis, then, certifies that, in terms of local content, the present bounds allow a higher degree of nonlocal correlations than those reported in [ADR82, WJS$^+$98, KSW$^+$05, ZYC$^+$03, PBS$^+$11b] or in any other previous experiment of our knowledge.

4.2 Multipartite nonlocal correlations suitable for device-independent information protocols

As extensively studied in the previous section, given some correlations between the measurement results on two parts, the local fraction [EPR92] quantifies the number of events that can be described by a local model. As such, it can be taken as a measure of nonlocality. Apart from maximal nonlocality, another extreme property of correlations is that of *monogamy* with respect to general nonsignaling correlations. Any given (nonsignaling) N-partite correlations are monogamous if the only nonsignaling extension of them to $N+1$ parts is the trivial one in which the part $N+1$ is uncorrelated to the initial N parts. Monogamy of correlations is clearly a very desirable property for cryptographic purposes. Note, however, that local deterministic correlations are monogamous but useless for cryptography. This is where the third ingredient comes into play: *randomness*. The correlations have to be such that the local outcomes are fully unpredictable by an adversary. A nonlocal fraction of unity is necessary but not sufficient both for the monogamy and full randomness of nonlocal correlations.

In Ref. [BKP06], Barrett, Kent, and Pironio showed that bipartite maximally entangled states can yield maximally nonlocal and monogamous correlations with fully random outcomes. They first exploited the fact that these states maximally violate the chained inequality [BC90], which implies that the nonlocal fraction is one. Then, contrary to other examples of bipartite maximally nonlocal correlations [HR83], they proved that the correlations leading to the maximal violation of the chained inequality also have the properties of being monogamous and having fully random local outcomes.

In a general multipartite scenario with N parts, these questions have hardly been considered (see, however, [ACSA10]). The multipartite situation is conceptually richer, as apart from the bare division between local and nonlocal, correlations allow for finer subclassifications in terms of locality among the different partitions. Indeed, one can consider k-local models in which the N parts are split into $k < N$ groups such that (i) the parties within each group can make use of any nonlocal resource, but (ii) the k groups are only classically correlated. Any correlations that can be reproduced by these models do not contain *genuine-multipartite nonlocality*, as nonlocal resources among only subsets of the N parts suffice. As in the case of locality in the bipartite setting, it is possible to construct inequalities to detect genuine-multipartite nonlocality, known as Svetlichny inequalities [Sve87]. A maximal violation of a Svetlichny inequality implies that the corresponding

correlations are maximally genuinely multipartite nonlocal.

It was an open question whether there exist fully genuinely multipartite nonlocal correlations with a quantum realization [ACSA10]. We show here that this is the case: Fully genuine-multipartite nonlocal correlations can be derived from Greenberger-Horne-Zeilinger (GHZ) states [GHZ89] of any number N of parts and local dimension d. To this end, we construct a family of Svetlichny inequalities generalizing the bipartite chained inequality, and show that GHZ states attain the algebraic violation in the limit of an infinite number of measurements. Then, we prove that the corresponding nonlocal quantum correlations are monogamous and fully random in the sense that the outcomes of any choice of $m < N$ parts provides m perfect random dits. Finally, we draw some implications on device-independent secret sharing.

Before proceeding, it is worth mentioning the relation between our results and Ref. [ACSA10]. There, criteria for the detection of quantum states with maximally genuine-multipartite correlations were provided. Using these criteria, it was shown that all graph states lead to such extremal nonlocal correlations. However, there genuine-multipartite nonlocality was studied with respect to k-local models in which the correlations among parties within each of the k groups are, for each value of the hidden variable, nonsignaling. In contrast, the k local models considered here are the most general ones, as no constraint is imposed on the correlations among parties within the same group. In this fully general scenario, no example of fully genuine-multipartite nonlocal correlations with quantum realization was known. Moreover, monogamy and randomness in a general multipartite scenario had not been considered previously either.

4.2.1 Background

Genunine multipartite nonlocality and bi-local content

As explained in preliminaries, there exist a notion a genuine multipartite nonlocality, analogous to the genuine multipartite entanglement displayed by quantum states. Consider for the sake of simplicity a scenario with three observers obtaining a probability distribution $P(A, B, C|X, Y, Z)$. We recall from section 2.2.3 that the probability distribution is said to be genuine multipartite nonlocal if it cannot be written as

$$\begin{aligned} P(A,B,C|X,Y,Z) &= q_{A|BC}\, P_{A|BC}(A,B,C|X,Y,Z) \\ &+ q_{B|AC}\, P_{B|AC}(A,B,C|X,Y,Z) \\ &+ q_{C|AB}\, P_{C|AB}(A,B,C|X,Y,Z) \end{aligned} \tag{4.8}$$

with $q_{A|BC}+q_{B|AC}+q_{C|AB}=1$, $q_{A|BC}, q_{B|AC}, q_{C|AB} \geq 0$, and $P_{A_1|A_2A_3}$ being a probability distribution local along the bipartition $A_1|A_2A_3$, (*i.e.* $P_{A_1|A_2A_3} = \sum_\lambda p(\lambda) P_{A_1} P_{A_2A_3}$).

Equivalently as the local content for the case of two observers, one can define a bi-local content for the case of three observers. The bi-local content is defined as the fraction of events that admit a description which is local along at least one of the bipartitions.

Definition 4.3. *Consider a non-signaling probability distribution among three parties $P(A,B,C|X,Y,Z)$. The bi-local content of $P(A,B,C|X,Y,Z)$, denoted by p_{BL}, is defined as*

$$\begin{aligned} p_{BL} &= \max_{P_{A|BC}, P_{B|AC}, P_{C|AB}} q_{A|BC} + q_{B|AC} + q_{C|AB} \\ \textit{such that } P &= q_{A|BC}\, P_{A|BC} + q_{B|AC}\, P_{B|AC} + q_{C|AB}\, P_{C|AB} \\ &\quad + (1 - q_{A|BC} + q_{B|AC} + q_{C|AB}) P_{NS} \end{aligned} \tag{4.9}$$

where P_{NS} is an arbitrary non-signaling probability distribution.

This construction generalizes straightforwardly to the N-partite scenario: the bi-local content is the fraction of events that admit a description which is local along any bipartition.

Equivalently to the bipartite case, Bell inequalities can be used to bound the bi-local content of a probability distribution. One has to consider Bell-inequalities that cannot be violated by probability distributions fulfilling (4.8). These are the so called Svetlichny-like inequalities.

Definition 4.4. *A Svetlichny inequality is a linear combination of the set of probabilities $C \cdot P = \sum C^{x,y,z}_{a,b,c} P(a,b,c|x,y,z)$ such that: (i) $C \cdot P \leq C_{BL}$ for all $P(A,B,C|X,Y,Z)$ fulfilling (4.8) and C_{BL} being a real constant, (ii) there exist a probability distribution $\tilde{P}(A,B,C|X,Y,Z)$ such that $C \cdot \tilde{P} > C_{BL}$. The probability distribution $\tilde{P}$ is then genuine multipartite nonlocal.*

Equivalently to the case of Bell inequalities and local content for the bipartite case, one can easily check that if $C \cdot \tilde{P} = \max_{P_{\rm NS} \in \mathcal{P}} C \cdot P_{\rm NS}$, then the bi-local of $\tilde{P}$ content is zero. Such correlations are called maximally genuine nonlocal.

It was an open question whether there exist quantum correlations with bi-local content equal to zero. Here we provide such correlations. Furthermore, we show that these correlations are monogamous and random, two properties that we explain now in detail and that make them appealing for device-independent multipartite protocols.

Monogamy and randomness

Consider an N-partite probability distribution $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$ that one uses to distribute secret information between the N parties in competition with a malicious party, the eavesdropper. Then, one should consider the $N+1$ extended probability distribution $P(A_1, \ldots, A_N, E|X_1, \ldots, X_N, Q)$. Clearly, a desired situation is to certify that the eavesdropper is completely uncorrelated with the statistics of the N parties. If that is the case, one says the the N-partite probability distribution is monogamous.

Definition 4.5. *The N-partite probability distribution $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$ is said to be monogamous if the only $N+1$-nonsignaling extension fulfills*

$$P(A_1, \ldots, A_N, E|X_1, \ldots, X_N, Q) = P(A_1, \ldots, A_N|X_1, \ldots, X_N)P(E|Q). \tag{4.10}$$

In Ref [MAG06] it is shown that $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$ being an extremal point of the nosignaling polytope ensures monogamy.

Nevertheless, monogamy is not sufficient to certify that the eavesdropper cannot acquire information about the secrecy distributed by the N parties. Another required ingredient is *local randomness*, *i.e.* that the outputs are equally likely. Otherwise the eavesdropper, however uncorrelated to the N parties, would hold some information about the secret just by betting for the most probable outcome.

Definition 4.6. *The N-partite probability distribution $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$ is fully locally random if for all $i \in \{1, \ldots, N\}$ the $N-1$-partite marginal probability distribution fulfills*

$$\begin{aligned} &P(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \\ &\qquad = \sum_{a_i} P(a_1, \ldots, a_N|x_1, \ldots, x_N) = \frac{1}{d^{N-1}} \end{aligned} \tag{4.11}$$

for all outputs and at least one input combination [1]*. Note that this is the strongest notion of local randomness, as it implies that any $N-k$ marginal is also random.*

[1] It is not necessary to demand this property for all the input combinations. One can perform the secrecy protocol whenever the chosen inputs are such they fulfill the randomness condition and discard the outputs otherwise.

A probability distribution that displays monogamy and local randomness is, roughly speaking, as good as possible for secrecy protocols. Such correlations have been found for the bipartite scenario using the chained Bell inequality as explained below.

Bipartite chained Bell inequality

Let us start by reviewing the Bell inequality used in the bipartite case to find maximally nonlocal, monogamous and locally random correlations. The bipartite chained Bell inequality for M settings and d outcomes can be expressed as [BKP06]

$$I_M^2 = \sum_{\alpha=1}^{M} \left(\langle [A_\alpha - B_\alpha]_d \rangle + \langle [B_\alpha - A_{\alpha+1}] \rangle \right) \geq d-1, \tag{4.12}$$

where $\langle \Omega \rangle$ stands for the average $\sum_{i=1}^{d-1} iP(\Omega = i)$, with

$$P(\Omega(A_x, B_y) = i) = \sum_{\{A_x, B_y\}/\Omega(A_x,B_y)=i} P(A_x, B_y|x, y) \tag{4.13}$$

the probability that random variable Ω_α is observed to take value i, $[\Omega_\alpha]$ is Ω_α modulo d, and $A_{M+1} = [A_1 + 1]$. Inequality (4.12) is satisfied by all local correlations and algebraically violated by the correlations of the maximally entangled state $|\Psi_d^2\rangle = \frac{1}{\sqrt{d}} \sum_{q=0}^{d-1} |qq\rangle$ [BKP06, CR08]. More precisely, measuring the quantum observables $\hat{A}_\alpha = \sum_{r_{A_\alpha}=0}^{d-1} r_{A_\alpha} |r_{A_\alpha}\rangle\langle r_{A_\alpha}|$ and $\hat{B}_\beta = \sum_{r_{B_\beta}=0}^{d-1} r_{B_\beta} |r_{B_\beta}\rangle\langle r_{B_\alpha}|$, where

$$\begin{aligned} |r_{A_\alpha}\rangle &= \frac{1}{\sqrt{d}} \sum_{q=0}^{d-1} e^{\frac{2\pi i}{d} q (r_{A_\alpha} - \frac{\alpha - 1/2}{M})} |q\rangle, \\ \text{and } |r_{B_\beta}\rangle &= \frac{1}{\sqrt{d}} \sum_{q=0}^{d-1} e^{-\frac{2\pi i}{d} q (r_{B_\beta} - \frac{\beta}{M})} |q\rangle, \end{aligned} \tag{4.14}$$

for $\alpha, \beta = 1, \ ..., \ M$, on $|\Psi_d^2\rangle$, leads to a Bell value that for large M can be well approximated as

$$I_M^2(\Psi_d^2) \approx \frac{\pi^2}{4d^2 M} \sum_{i=1}^{d-1} i / \sin^2\left(\frac{\pi i}{d}\right). \tag{4.15}$$

This value tends to 0 as M grows. Since all the terms in (4.12) are by definition non-negative, this is the maximal violation any probability distribution can render.

4.2.2 Extension to the multipartite case

Tripartite case

Let us now extend inequality (4.12) to the multipartite scenario. We first discuss the case $N = 3$ and extend the formalism to arbitrary N later. Consider then three random variables A_α, B_β, and C_γ, for $\alpha, \beta, \gamma = 1, \ldots, M$, each of d possible outcomes $\{0, \ldots, d-1\}$, measured by Alice, Bob, and Charlie, respectively.

Theorem 4.7. *The inequality*

$$I_M^3 = \sum_{\alpha,\beta=1}^{M} \left(\langle [A_\alpha - B_{\alpha+\beta-1} + C_\beta] \rangle + \langle [B_{\alpha+\beta-1} - A_{\alpha+1} - C_\beta] \rangle \right) \geq M(d-1) \tag{4.16}$$

is fulfilled by all the bi-local probability distributions with decomposition of the form (4.8)*. Furthermore, it is maximally violated by quantum correlations. Hence, those quantum correlations are maximally genuine nonlocal.*

Proof. Here we have introduced $B_{M+\nu} = [B_\nu + 1]$, for any $\nu = 1, \ldots, M$. Given that (4.12) is a bipartite Bell inequality, the fact that the tripartite inequality is fulfilled by all correlations local in any bipartition can be seen with an argument similar to one of the arguments of [BBGL11]. The local relabeling $B_\alpha \to B_{\alpha+\beta-1}$ of Bob's bases in I_M^2 gives $I_M^2(\beta) = \sum_{\alpha=1}^{M} \left(\langle [A_\alpha - B_{\alpha+\beta-1}] \rangle + \langle [B_{\alpha+\beta-1} - A_{\alpha+1}] \rangle \right)$. Since this simply defines a symmetry[2] of (4.12) it also fulfills the inequality $I_M^2(\beta) \geq d-1$. In turn, the β-th term in the definition of I_M^3 can be recast as $I_M^2(\beta) \circ C_\beta = \sum_{\alpha=1}^{M} \left(\langle [A_\alpha - B_{\alpha+\beta-1} - C_\beta] \rangle + \langle [B_{\alpha+\beta-1} - A_{\alpha+1} - C_\beta] \rangle \right)$, where "$\circ\, C_\beta$" stands for the "insertion of C_β with the opposite sign from $B_{\alpha+\beta-1}$." Grouping Bob and Charlie together with a single effective variable $B_{\alpha+\beta-1} - C_\beta$ we see that, for any correlations local with respect to the bipartition $A : BC$, it must be $I_M^2(\beta) \circ C_\beta \geq d-1$. In addition, since this holds for all β and $I_M^3 \equiv \sum_{\beta=1}^{M} I_M^2(\beta) \circ C_\beta$, any correlations local with respect to $A : BC$ satisfy $I_M^3 \geq M(d-1)$. The same reasoning holds of course for correlations local with respect to $B : AC$ and an effective variable $A_\alpha + C_\beta$. Finally, since I_M^3 is symmetric with respect to the permutation of A and C (see Appendix A.2), the tripartite inequality must be satisfied by *all* probability distributions with a bilocal model. That

[2] A symmetry in the sense that a relabeling of measurements or outcomes leads to the original inequality. As relabeling is a pure classical operation that can be performed locally, hence a symmetry of a Bell inequality is also a Bell inequality with the same properties.

is, by all the distributions that can be written as a convex combination of correlations with a local model with respect to any bipartition of the three parties. For later convenience, instead of (4.16), we consider its regularized version $\overline{I_M^3} = I_M^3/M$:

$$\overline{I_M^3} = \frac{1}{M}\sum_{\alpha,\beta=1}^{M}\left(\langle[A_\alpha - B_{\alpha+\beta-1} + C_\beta]\rangle + \langle[B_{\alpha+\beta-1} - A_{\alpha+1} - C_\beta]\rangle\right) \geq d-1. \tag{4.17}$$

For the quantum realization, we introduce first Charlie's observable $\hat{C}_\gamma = \sum_{r_{C_\gamma}=0}^{d-1} r_{C_\gamma}|r_{C_\gamma}\rangle\langle r_{C_\gamma}|$, where

$$|r_{C_\gamma}\rangle = \frac{1}{\sqrt{d}}\sum_{q=0}^{d-1} e^{\frac{2\pi i}{d}q(r_{C_\gamma} - \frac{\gamma-1}{M})}|q\rangle, \tag{4.18}$$

for $\gamma = 1,\ ...,\ M$. Measurements are performed on the d-dimensinoal GHZ state $|\psi_d^3\rangle = \sum_{q=0}^{d-1}|qqq\rangle$. In appendix A.3 it is shown that $I_M^2(\Psi_d^2) = \overline{I_M^3}(\Psi_d^3) = 0$. Thus, it displays the maximal nosignaling violation of inequality (4.17).

□

Arbitrary number of parties

For arbitrary N, the inequality generalizes as $\overline{I_M^N} = \frac{1}{M}\sum_{\psi=1}^{M}\overline{I_M^{N-1}}(\psi)\circ Z_\psi$, where Z is the N-th variable and the generalization followed by induction. This gives

$$\begin{aligned}\overline{I_M^N} &= \frac{1}{M^{N-2}}\sum_{\alpha,\beta,\ldots,\chi,\psi=1}^{M}\left(\langle[A_\alpha - B_{\alpha+\beta-1} + ... - (-1)^{N-1}Y_{\chi+\psi-1} + (-1)^{N-1}Z_\psi]\rangle\right.\\ &+ \left.\langle[B_{\alpha+\beta-1} - A_{\alpha+1} - ... + (-1)^{N-1}Y_{\chi+\psi-1} - (-1)^{N-1}Z_\psi]\rangle\right) \geq d-1,\end{aligned} \tag{4.19}$$

for N random variables A_α, B_β, C_γ, ..., Y_ψ, and Z_ζ, in possession of Alice, Bob, Charlie, ..., Yakira, and Zack, respectively. Here we introduce this alphabetic notation instead of the common $A_1, A_2, \ldots A_N$ in order to avoid an overpopulation of indexes. Also, for $\Omega = A, B, C, ..., Y$, or Z, we have introduced, in a general way, $\Omega_{i\times M+\omega} = [\Omega_\omega + i]$, for any integer i and all $\omega = 1,\ ...,\ M$. In the generic case of arbitrary N, the composition rule "$\circ$" refers to "insertion of the new variable with the opposite sign from the previously inserted one." With the same reasoning [BBGL11] as above, from the fact that $\overline{I_M^{N-1}} \geq d-1$ is satisfied by all $(N-1)$-partite correlations local in at least one bipartition, it follows by construction that (4.19) is

satisfied by any N-partite correlations local with respect to any bipartition "Z with at least anyone else versus the rest." Once again, by symmetry under the permutation of Z with some other part (see Appendix A.2), one sees that (4.19) is satisfied by all N-partite distributions local in a bipartition. Equation (4.19) is the Svetlichny inequality used in what follows to prove our results. Actually, equation (4.19) encapsulates an entire family of Svetlichny inequalities for M measurements of d outcomes. The same family of inequalities is independently derived in [BBB+12].

In order to show that there exist a quantum realization of correlations that violate maximally (4.19), we introduce analogous observables to the one for the tripartite case (4.18). For instance, for Yakira and Zack we define $\hat{Y}_\psi$ and $\hat{Z}_\zeta$ respectively with eigenstates

$$\begin{aligned} |r_{Y_\psi}\rangle &= \frac{1}{\sqrt{d}}\sum_{q=0}^{d-1} e^{-(-1)^{N-1}\frac{2\pi i}{d}q(r_{Y_\psi}-\frac{\psi-1}{M})}|q\rangle \\ \text{and } |r_{Z_\zeta}\rangle &= \frac{1}{\sqrt{d}}\sum_{q=0}^{d-1} e^{(-1)^{N-1}\frac{2\pi i}{d}q(r_{\hat{Z}_\zeta}-\frac{\zeta-1}{M})}|q\rangle, \end{aligned} \tag{4.20}$$

for $\psi, \zeta = 1,\ ...,\ M$. In the limit $M \to \infty$ the maximal violation of inequality (4.19) is obtained by measuring these observables on the N-partite GHZ state $|\Psi_d^N\rangle = \frac{1}{\sqrt{d}}\sum_{q=0}^{d-1}|qqq\ ...\ qq\rangle$. To see this we show in Appendix A.3 that

$$\overline{I_M^2}(\Psi_d^2) = \overline{I_M^3}(\Psi_d^3) = ... = \overline{I_M^N}(\Psi_d^N), \tag{4.21}$$

where $\overline{I_M^3}(\Psi_d^3)$ and $\overline{I_M^N}(\Psi_d^N)$ are, respectively, the Bell values of (4.17) and (4.19) for the observables defined above on states $|\Psi_d^3\rangle$ and $|\Psi_d^N\rangle$. Thus, the Bell values for all N equally tend to zero as M grows. Since inequality (4.19) consists, as in the bipartite case, exclusively of non-negative terms, in the limit $M \to \infty$ GHZ states attain its algebraic violation. Furthermore, since the inequality is only violated by genuinely multipartite nonlocal correlations, the latter implies that all GHZ states are maximally genuine-multipartite nonlocal.

4.2.3 Monogamy and randomness

Any correlations P featuring $\overline{I_M^N}(P) = 0$ must necessarily satisfy

$$P(r_A \neq [r_B - ... - (-1)^{N-1}r_Z], r_B, ..., r_Z|\alpha, \beta, ..., \zeta) \equiv 0, \tag{4.22}$$

for all $r_A,\ r_B,\ ...,\ r_Z \in \{0, 1, ..., d-1\}$ and $(\alpha, \beta, ..., \zeta)$ being any of the $2M^{N-1}$ measurement bases appearing in (4.19). Note that not all possible

combinations of the M local bases appear in the inequality. Hereafter, we study the properties of the probability distributions P corresponding to the bases appearing in (4.19). For each such distribution, $d^N - d^{N-1}$ coefficients are automatically set to zero by (4.22). The remaining d^{N-1} are univocally determined by the marginal probability distribution corresponding to any $N-1$ parts. For instance,

$$\begin{aligned} & \sum_{r_A} P(r_A, r_B, ..., r_Z|\alpha, \beta, ..., \zeta) \\ = \; & P(r_A = [r_B - ... - (-1)^{N-1} r_Z], r_B, ..., r_Z|\alpha, \beta, ..., \zeta) \\ = \; & P(r_B, ..., r_Z|\beta, ..., \zeta), \end{aligned} \tag{4.23}$$

and equivalently for other parties and measurement bases. When $M \to \infty$, the following theorem fixes the value of all $(N-1)$-partite marginals and hence imposes uniqueness. In turn, the uniqueness of P implies also its monogamy [MAG06]. Moreover, the theorem proves also the full randomness of all its marginal distributions.

Theorem 4.8. *For any N-partite nonsignaling distribution P such that $\overline{I_M^N}(P) \leq \varepsilon$, with $\varepsilon \geq 0$, the marginal distributions fulfill*

$$P\big(\mathcal{S}(r_A, ..., r_Z)|\mathcal{S}(\alpha, ..., \zeta)\big) \leq \frac{1}{d^{N-1}} + \frac{d(N-1)}{4}\varepsilon, \tag{4.24}$$

for $(\alpha, ..., \zeta)$ any of the settings appearing in (4.19)*, where $\mathcal{S}$ refers to any subset of $N-1$ parts out of all N.*

The proof of the theorem is provided in A.1. Note that for P realized by GHZ states $|\Psi_d^N\rangle$ and the measurements considered here, it is $\overline{I_M^N}(P) \approx \frac{\pi^2}{4d^2M}\sum_{i=1}^{d-1} i/\sin^2(\frac{\pi i}{d})$, which tends to 0 as M grows. Therefore, the GHZ-state quantum realization fulfills the theorem for any arbitrarily small ε. In this limit, the theorem thus guarantees that $P\big(\mathcal{S}(r_A, ..., r_Z)|\mathcal{S}(\alpha, ..., \zeta)\big) = \frac{1}{d^{N-1}}$. That is, that all the $(N-1)$-partite marginal distributions (and therefore all the marginal distributions) have each and all of their outcomes equally probable, or in other words, that they are fully random.

4.2.4 Device-independent secret sharing

Monogamy of multipartite correlations is a desired property in multipartite cryptographic scenarios. In particular, for instance, if $\overline{I_M^N}(P) = 0$ then correlations P fulfill the requirements for a device-independent implementation of the quantum secret-sharing protocol introduced in [HBB99]. We analyze

this for the particular case $N = 3$ for ease of notation, but the same conclusions are valid for any $N \geq 3$. Alice wishes to share secret dits with Bob and Charlie, but she suspects that one of them is dishonest. Therefore, she wishes to do it in such a way that Bob and Charlie can access the value of the dits only if they are together. The three distant users then randomly input settings α, β, and γ into three black boxes described by correlations $P(r_A, r_B, r_C|\alpha, \beta, \gamma)$ with the property that $\overline{I_M^3}(P) = 0$. They repeat the procedure many times, each time recording the outcome, and at the end publicly broadcast all the settings used. From Theorem 1 they know that whenever their settings happen to match those of (4.17), i.e., $\alpha - \beta + \gamma - 1 = 0$ (modulo M), or $\alpha - \beta + \gamma = 0$ (modulo M), then $P(r_A = a|[r_C - r_B] = a) \equiv 1$ for all $a \in \{0, 1, ..., d-1\}$. This means that then, if Bob and Charlie meet, they can determine with certainty the value of Alice's dit r_A simply by subtracting their outcomes. In addition, since all marginals of P (for the relevant settings) are fully random, neither Bob nor Charlie can obtain any information at all from their local outcomes alone. Finally, as the correlations are monogamous, Alice's dit is also unpredictable by any external adversary.

4.2.5 Conclusions and discussion

We presented a multipartite version of the multiple-setting multiple-outcome chained Bell inequalities. The inequalities introduced are Svetlichny-like: they are satisfied by all probability distributions expressed as mixtures of local correlations with respect to any bipartition. We showed that, in the limit of an infinite number of settings, correlations from GHZ states of any local dimensions or numbers of parts violate these inequalities as much as any non-signaling correlations. This proves that the genuine-multipartite nonlocal content of GHZ states is maximal. Moreover, we showed that any correlations algebraically violating the present inequalities are monogamous with respect to nonsignaling compositions and yield fully random outcomes for any subset of parts. This proves monogamy and full randomness of genuinely multipartite quantum correlations in a nonsignaling scenario. Finally, we showed that the correlations from GHZ states approach, as the number of measurement settings grows, those required for device-independent secret sharing secure against eavesdroppers limited solely by the no-signaling principle.

Chapter 5

An operational framework for quantum correlations

It is often useful, when defining a framework for a new field, to establish analogies with other fields that have already gone through this process. This allows one to identify the elements that the new field lacks in comparison with the elder and to find hints for its characterization. This situation matches perfectly the case of nonlocality and entanglement. The former, however first established by Bell's theorem in the early sixties, has only been recently established as an important resource to perform quantum information protocols. On the other hand, entanglement has been already extensively studied and characterized.

In particular we focus on the operational framework that is used to study entanglement as a resource. Local Operations assisted by Classical Communication (LOCC) play a key role in this framework, and indeed entanglement can simply be defined as a resource that cannot be created by LOCC. This operational definition is consistent with other mathematical criteria expressed in terms of the structure of the quantum state itself and that provide alternative definitions of entanglement. The generalization to multipartite scenarios is straightforward: genuine multipartite entanglement is a resource that cannot be created by LOCC, even if one allows a subset of parties to collaborate. Again, this operational definition is consistent with mathematical criteria based on the quantum state that define multipartite entanglement.

Surprisingly, an analogous operational definition for nonlocality has been neither established nor studied. In the second section of this chapter we provide a natural operational definition of nonlocality and we compare it with

the canonical definitions of nonlocality based on Bell local decompositions. Remarkably, we find that the standard definition of genuine multipartite nonlocality is inconsistent with the operational framework. We provide alternatives to recover consistency and discuss their main features.

Throughout the third section we apply the tools provided by the operational framework for nonlocality to the study of information theoretic principles for quantum mechanics. Roughly speaking, information theoretic principles aim at defining the set of quantum correlations among distant observers without making any reference to the structure of Hilbert spaces. Two particular principles are considered potential good candidates: Information causality and Non-triviality of communication complexity. Here, we show that both are insufficient and furthermore, we recognize the features that such principle should display in order to define quantum correlations.

5.1 An operational framework for nonlocality

Let us first review the most important features of the well-known operational framework for entanglement theory. The formalism for nonlocality will be established thereafter, with special focus on the analogies and differences between the two.

5.1.1 The framework of LOCC

The first step when deriving this theoretical formalism consists in identifying the relevant objects and set of operations, see Table 5.1. The relevant objects in the entanglement scenario are quantum states in systems composed by N observers, labeled by A_i with $i = 1, \ldots, N$. The relevant set of operations is the set of LOCC. The whole formalism then relies on the following principle, which has a clear operational motivation: *entanglement of a quantum state is a resource that cannot be created by LOCC.* This implies that those states that can be created by LOCC are not entangled. These states are called separable and can be written as [HHHH09]

$$\rho_{A_1 \ldots A_N} = \sum_j p_j \rho^j_{A_1} \otimes \cdots \otimes \rho^j_{A_N}. \tag{5.1}$$

In turn, those states that cannot be created by LOCC are entangled and require a nonlocal quantum resource for the preparation. It is easy to see that LOCC protocols map separable states into separable states. Finally, entanglement witnesses are Hermitian operators W such that (i) $\mathrm{Tr}(W\rho_S) \geq$

Resource	Objects	Operations
Entanglement	Quantum states	LOCC
Nonlocality	Joint Probability Distributions	WCCPI

Table 5.1: Comparison of entanglement and non-locality from an operational point of view. Once the basic ingredients of the theory have been identified, an operational framework is based on the following principle: the resource contained in the states cannot increase under the set of operations.

0 for all separable states ρ_S but (ii) there exist an entangled state ρ such that $\mathrm{Tr}(W\rho) < 0$.

The picture becomes richer when considering intermediate cases where only some of the N parties share entangled states. For simplicity we restrict our considerations in what follows to three parties. Consider an entangled state in which only two parties, say A_2 and A_3 are entangled. The corresponding state is called biseparable and can be written as

$$\rho_{A_1A_2A_3} = \sum_j p_j \rho^j_{A_1} \otimes \rho^j_{A_2A_3}. \tag{5.2}$$

This state is not genuine 3-partite entangled, as for its LOCC creation, it suffices that two of the parties act together. Similarly as above, (i) LOCC protocols where A_2 and A_3 act together map biseparable states into biseparable states along the bipartition $A_1 - A_2A_3$ and (ii) these states do not violate any entanglement witness along this partition. Finally, it is possible to define entanglement witness for genuine 3-partite entanglement, which are positive when acting on biseparable states along any bipartition, but give a negative value for some states.

5.1.2 The formalism of WCCPI

As extensively introduced and discussed in chapter 2, within the device-independent formalism the relevant objects are the sets of joint probability distributions. That is, consider N distant observers. Each observer i can input a classical variable x_i in his system, which produces a classical outpu a_i. The correlations among the input/output processes in each system are described by the joint probability distribution $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$. The main reason why device-independent applications are possible in the quantum regime is because of the existence of nonlocal quantum correlations. Therefore it is nonlocality the resource that we characterize operationally in the following.

In a similar way as it is done for entanglement, the first step consists in identifying the relevant objects and set of operations, see Table 5.1. The relevant objects are the joint probability distributions $P(A_1, \ldots, A_N|X_1, \ldots, X_N)$. The corresponding set of operations should include local processing of the classical inputs and outputs. These operations, by definition, do not create nonlocality. On the other hand, communication is allowed only if it takes place before the inputs are known, otherwise it can be used to create nonlocal correlations. Such communication taking place before the inputs are known can be used either to generate shared randomness or to announce the outcomes of a sequence of measurements prior to the realization of the nonlocal experiment. A general protocol in the nonlocality scenario would thus begin with a *preparation phase*, where one of the parties would measure its system and broadcast the measurement outcome. On the basis of that result, a second party measures its system, etc. At the end of the preparation phase, the parties exchange some shared randomness and announce that they are ready for the nonlocal experiment. Communication between them is forbidden from this point on. The second step is the *measurement phase*, where each party is given an input, or question, and they compute the outcome or answer by using the correlations resulting from the *preparation phase* and by processing the obtained classical information at will. The last process is commonly referred to as 'wirings'. Thus, in the nonlocality framework, the set of relevant operations is Wirings & Classical Communication Prior to the Inputs (WCCPI).

Once these two ingredients are identified, it is straightforward to obtain an operational definition of nonlocality: *nonlocality of correlations* $P(a_1, \ldots, a_N|x_1, \ldots, x_N)$ *is a resource that cannot be created by WCCPI.*

Not surprisingly, this operational definition leads to the standard definition of nonlocality due to Bell [Bel64] when considering N distant parties. Indeed, it is easy to see that the correlations that can be created by WCCPI have the form, see Eq. (5.1),

$$P_{\mathrm{L}}(a_1, \ldots, a_n|x_1, \ldots, x_N) = \sum_{\lambda} p(\lambda) P_1(a_1|x_1, \lambda) \ldots P_1(a_n|x_n, \lambda), \qquad (5.3)$$

in which the local maps $P_i(a_i|x_i, \lambda)$ produce the classical output a_i depending on the input x_i and a shared classical random variable λ. All correlations that admit a decomposition (5.3) are local, while they are nonlocal otherwise. WCCPI protocols map local correlations into local correlations. Finally, nonlocality can be detected by the violation of Bell inequalities (see chapter 2) such that, with $\vec{c}$ containing the coefficients of the Bell inequality (i)

$C \cdot P_{\mathrm{L}} \geq 0$ for all local correlations P_{L} but (ii) there exist correlations P such that $C \cdot P < 0$.

As for entanglement, the next step is to characterize genuine multipartite nonlocality. This question has already been studied and the standard definition of genuine multipartite nonlocalty is due to Svetlichny [Sve87]. We restrict our considerations again to three parties and the partition $A_1 - A_2A_3$ for sake of simplicty. According to Svetlichny, correlations that can be written as, see Eq. (5.2),

$$P(a_1, a_2, a_3|x_1, x_2, x_3) = \sum_{\lambda} p(\lambda) P_1(a_1|x_1, \lambda) P_{23}(a_2, a_3|x_2, x_3, \lambda) \tag{5.4}$$

do not contain any genuine tripartite nonlocality, as there is a local decomposition when parties A_2 and A_3 are together. Correlations admitting a decomposition like (5.4) are named in what follows bilocal (BL). As it happened for entanglement and LOCC, it is expected that under WCCPI protocols along the partition $A_1 - A_2A_3$, bilocal correlations are mapped into bilocal correlations. Consequently, no bipartite Bell inequality between A_1 and A_2A_3 can be violated. Remarkably, we prove here that this intuition is incorrect. This implies that the standard definition of genuine multipartite nonlocality, given by (5.4), is inconsistent with the operational approach. In the following we show examples of the inconsistencies and also provide and discuss alternative definitions of genuine multipartite nonlocality that are consistent with our operational framework.

5.1.3 Inconsistencies of bilocal decompositions with the operational formalism

We show the inconsistencies of the definition of BL by providing correlations that (i) have a decomposition of the form (5.4) and (ii) become nonlocal along the partition $A_1 - A_2A_3$ when a WCCPI protocol, where A_2 and A_3 collaborate, is implemented. An example of these correlations with a quantum realization can be established in the simplest scenario consisting of two measurements of two outcomes for each of the three observers. The measurements are performed on the quantum state $|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$. The first observer, A_1 measures σ_z and σ_x, labeled by $x_1 = 0$ and $x_1 = 1$ respectively. A_2 measures σ_z and σ_x, labeled by $x_2 = 0$ and $x_2 = 1$ respectively. A_3 measures $\frac{\sigma_z+\sigma_x}{\sqrt{2}}$ and $\frac{\sigma_z-\sigma_x}{\sqrt{2}}$, labeled by $x_3 = 0$ and $x_3 = 1$ respectively. These correlations have a decomposition of the form (5.4). This can be easily computed by a linear program, see Ref. [PBS11a].

Let us now see how these tripartite collaborations can be mapped into nonlocal bipartite correlations along the partition $A_1 - A_2A_3$ with an WCCPI protocol in which A_2 and A_3 collaborate. The protocol works as follows (see also Figure 5.1.b): the first observer A_1 obtains the output by using trivially his share of the tripartite box. A_2 and A_3 collaborate by using the output obtained by A_2 as input for A_3. The resulting tripartite probability distribution reads $P(A_1, A_2, A_3|X_1, X_2, X_3 = A_2)$. The final output is A_3's output, so that the final probability distribution is $P(A_1, A_3|X_1, X_2) = \sum_{a_2} P(A_1, a_2, A_3|X_1, X_2, A_2)$. This bipartite probability distribution $P(A_1, A_3|X_1, X_2)$ does not have a local model. This can be verified by calculating the value of the Clauser-Horne-Shimony-Holt (CHSH) polynomial

$$\beta = C(0,0) + C(0,1) + C(1,1) - C(1,0), \tag{5.5}$$

with $C(X_1, X_2) = P(a_1 = a_3|X_1, X_2) - P(a_1 \neq a_3|X_1, X_2)$. The value obtained is $\beta = \frac{3}{\sqrt{2}} \approx 2.12$. Local correlations fulfill $\beta \leq 2$, thus we conclude that the correlations are nonlocal along the partition $A_1 - A_2A_3$.

Alternatively, one can assess the inconsistency of Svetlichny's definition by noting that our tripartite example behaves non-locally if one of the parties broadcasts its measurement outcomes before the nonlocal experiment takes place. Indeed, suppose that, prior to the experiment, A_2 measures $x_2 = 1$. If the result is $a_2 = 1$, A_1 and A_3 are projected onto the distribution $P'(A_1, A_3|X_1, X_3) \equiv P(A_1, A_3|X_1, X_3, x_2 = 1, a_2 = 1)$. On the contrary, if A_2 reads $a_2 = -1$, A_1 receives the order of inverting her measurement outcomes for measurement $x_1 = 1$, and so the system is projected again into $P'(A_1, A_3|X_1, X_3) = P(-A_1, A_3|X_1, X_3, x_2 = 1, a_2 = -1)$. It can be checked that the new bipartite distribution $P'(A_1, A_3|X_1, X_3)$ violates the CHSH inequality maximally ($\beta = 2\sqrt{2} \approx 2.82$).

5.1.4 A new definition of multipartite nonlocality consistent with WCCPI

As mentioned above, the existence of these correlations implies that the standard definition of genuine multipartite nonlocality (5.4) is inconsistent with our operational approach, as it would imply that genuine tripartite nonlocality could be created by WCCPI when two parties collaborate [1]. Thus,

[1] Actually, the last construction suggests another venue to generate non-locality, namely, to make use of *Stochastic Wirings & Classical Communication Prior to the Inputs* (SWCCPI). However, it is easy to see that, whenever non-locality can be generated probabilistically with SWCCPI protocols, it can be also activated deterministically via WCCPI. Imagine, for instance, that $P_1(a_1, a_3|x_1, x_3) \equiv P(a_1, a_3|x_1, x_3, x_2 = 1, a_2 = +1)$ is non-

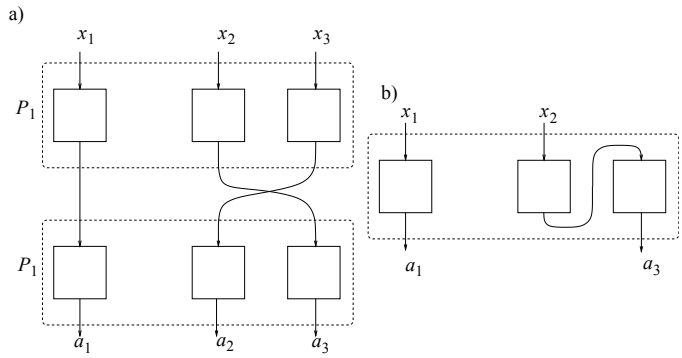

Figure 5.1: WCCPI protocols. a) Example of a WCCPI protocol consistent with the definition of genuine nonlocality provided by Eq. (5.4) (BL). The resulting probability distribution $P(a_1, a_2, a_3|x_1, x_2, x_3)$ does not violate any Bell inequality along the bipartition $A_1 - A_2A_3$. However, this may be no longer true for WCCPI protocols where inputs depend on outputs produced by the collaborating parties. b) WCCPI protocol in which the input of party A_3 is the output of A_2. As shown in the text, this protocol can map tripartite probability distributions with a bilocal decomposition as in (5.4) into bipartite distributions $P(a_1, a_3|x_1, x_2)$ that violate a Bell inequality. This proves that this type of WCCPI protocols is not compatible with the definition of genuine multipartite nonlocality (5.4).

the concept of genuine multipartite nonlocality is not correctly captured by Eq. (5.4).

Now, the natural question is whether there are definitions of bilocality which do not suffer from these inconsistencies. Or in other words, whether one can find an analogous version of equation (5.2) in the context of nonlocality consistent with the operational framework established by WCCPI. Before moving into that, it is worth understanding why the previous Bell violation is possible even if the correlations seem to have a proper local decomposition. The main reason is that no structure is imposed on the joint terms $P_{23}(A_2, A_3|X_2, X_3, \lambda)$; in particular, these terms may be incompatible with the no-signaling principle, e.g. $P_{23}(a_2|x_2, x_3, \lambda) \neq P_{23}(a_2|x_2, x_3', \lambda)$ for some λ. Further, if no structure is imposed on $P_{23}(A_2, A_3|X_2, X_3, \lambda)$ the decomposition (5.4) may include terms which display both signaling from A_2 to A_3 and from A_3 to A_2; that is, the outcome probability distribution of A_2 depends on A_3's input and viceversa. Hence, a decomposition including

local, but $P_2(a_1, a_3|x_1, x_3) \equiv P(a_1, a_3|x_1, x_3, x_2 = 1, a_2 = -1)$ is not. Let C be such that $C \cdot \vec{L} \geq 0$ for all bipartite local distributions $\vec{L}$ and $C \cdot \vec{P}_1 < 0$. In the event $a_2 = -1$, A_1 and A_3 receive the order of simulating a local box $P_2'(a_1, a_3|x_1, x_3)$ such that $\vec{c} \cdot \vec{P}_2 = 0$. Then, it is clear that the so-constructed box $Q(a_1, a_3|x_1, x_3) \equiv p(a_2 = 1|x_2)P_1(a_1, a_3|x_1, x_3) + p(a_2 = -1|x_2)P_2'(a_1, a_3|x_1, x_3)$ satisfies $C \cdot \vec{Q} < 0$, and, consequently, is non-local.

these terms cannot be considered a physical description of the situation in which one of the observers measures first. This is crucial in our previous example.

In our protocol the output of one of the parties is used as the input of the other party, or it is broadcast prior to the nonlocality experience. This implicitly assumes a temporal order in the measurements which is inconsistent with such decompositions. Indeed, all the examples of distributions of the form (5.4) leading to a Bell violation under WCCPI have to be such that the bilocal decomposition requires terms displaying signaling in both directions. Whether the converse is true, that is, whether every decomposition with such terms can be mapped via LOCC into a nonlocal one is an interesting open question. We come back to this point below.

It is now clear that tripartite correlations with bilocal models (5.4) such that all the terms $P_{23}(A_2, A_3|X_2, X_3, \lambda)$ satisfy the no-signaling principle, i.e. marginal distributions on A_2 (A_3) do not depend on the input by A_3 (A_2) for all λ, are consistent with our operational framework. We name these correlations no-signaling bilocal (NSBL). They are operationally understood as correlations obtained by collaborating parties sharing no-signaling resources. This definition however is too restrictive, as it excludes correlations obtained by protocols in which the collaborating parties communicate, which is perfectly valid within our framework. Indeed, we show next that NSBL correlations do not define the largest set of correlations compatible with our framework, see Figure 5.2.

Consider instead the set of tripartite no-signaling correlations (see appendix B.1 for the corresponding N-party generalization) that can be decomposed as

$$\begin{aligned} P(a_1a_2a_3|x_1x_2x_3) &= \sum_\lambda p_\lambda P(a_1|x_1,\lambda)P_{2\to 3}(a_2a_3|x_2x_3,\lambda) \\ &= \sum_\lambda p_\lambda P(a_1|x_1,\lambda)P_{2\leftarrow 3}(a_2a_3|x_2x_3,\lambda) \end{aligned} \tag{5.6}$$

with the distributions $P_{2\to 3}$ and $P_{2\leftarrow 3}$ obeying the conditions

$$P_{2\to 3}(a_2|x_2,\lambda) = \sum_{a_3} P_{2\to 3}(a_2a_3|x_2x_3,\lambda), \tag{5.7}$$

$$P_{2\leftarrow 3}(a_3|x_3,\lambda) = \sum_{a_2} P_{2\leftarrow 3}(a_2a_3|x_2x_3,\lambda). \tag{5.8}$$

We say that these correlations admit a *time-ordered bilocal (TOBL)* model in the bipartition $A_1 - A_2A_3$. As can be seen from relations (5.7) and (5.8) we

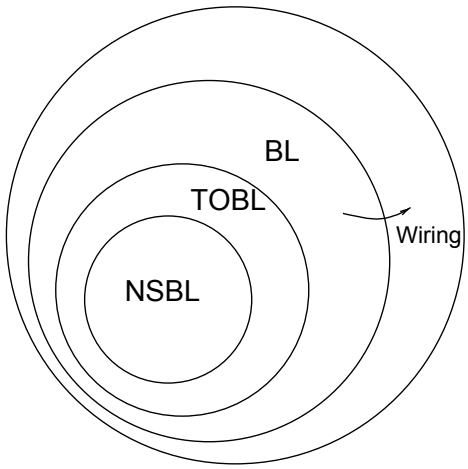

Figure 5.2: Representation of no-signaling, time-ordered and general bilocal correlations. We prove here that the set of non-signaling bilocal correlations (NSBL) is strictly contained in the set of time-ordered bilocal correlations (TOBL). The set TOBL is closed under wirings in the sense that LOCC protocols involving two collaborating parties, say A_2 and A_3, map TOBL correlations to correlations which are local with respect to the partition $A_1 - A_2A_3$. The set of general bilocal correlations (BL) however, contains correlations that can be mapped by LOCC protocols to correlations that violate a Bell inequality in the bipartition $A_1 - A_2A_3$.

impose the distributions $P_{2\to 3}$ and $P_{2\leftarrow 3}$ to allow for signaling at most in one direction, indicated by the arrow. Decomposition (5.6) has also been considered in [PBS11a], and has a clear operational meaning: $P(A_1A_2A_3|X_1X_2X_3)$ can be simulated by a classical random variable λ with probability distribution p_λ distributed between parts A_1 and the composite system A_2A_3. Using this variable, A_1 generates the output according to the distribution $P(A_1|X_1,\lambda)$; on the other side, the two outputs a_2, a_3 are generated using either $P_{2\to 3}(A_2A_3|X_2X_3,\lambda)$ or $P_{2\leftarrow 3}(A_2A_3|X_2X_3,\lambda)$, depending on which of the inputs x_2 or x_3 is used first.

We now show two results that show the importance of TOBL decompositions to characterize multipartite nonlocality. First, TOBL correlations are consistent with our operational point of view, as we show that any WCCPI protocol along the partition $A_1 - A_2A_3$ maps TOBL correlations (5.6) into probability distributions with a local model along this partition, this is our first theorem. On the other hand, one may wonder whether the set of TOBL correlations is indeed larger than the well-known NSBL set, which is itself also consistent with WCCPI protocols. This is precisely what is shown in our second theorem. These two results taken together show that

the set of TOBL correlations is a distinct set from the ones that have been previously considered in the literature, and furthermore, it is the adequate set to describe multipartite nonlocality.

Theorem 5.1. *Consider l tripartite systems behaving according to the probability distributions $P^1, \ldots, P^l$ respectively. If all the probability distributions have a TOBL decomposition along the bipartition $A_1 - A_2A_3$ (that is, a decompostion as in* (5.6)*), then any WCCPI along $A_1 - A_2A_3$ maps the l systems into a resultant system P_{fin} that is local along the same bipartition (that is, P_{fin} fulfills* (5.4)*).*

Proof. As it has been mentioned previously in the definition of WCCPI protocols, they can be divided into a preparation phase and a measurement phase. During the preparation phase, one can use communication to establish shared-randomness and/or measure on a subset of the l tripartite boxes and broadcast the outcome to the other observers. We first show that this measurement and outcome broadcasting process is indeed equivalent to shared randomness. Consider for example that one of the l tripartite boxes P^i is measured by the third observer, by inputing $\tilde{x}_3$ which provides an ouput $\tilde{a}_3$ which is broadcasted to the remaining parties. The other parties are then left with a bipartite system that behaves according to $P^i(A_1, A_2, |X_1, X_2, \tilde{x}_3, \tilde{a}_3)$. One can easily check that this probability distribution has a local model as

$$P(a_1a_2|x_1x_2\tilde{x}_3\tilde{a}_3) = \sum_\lambda p'_\lambda P(a_1|x_1, \lambda)P'(a_2|x_2, \lambda), \tag{5.9}$$

with

$$\begin{aligned} p'_\lambda &= \frac{p_\lambda}{P(\tilde{a}_3|\tilde{x}_3)} P_{2\leftarrow 3}(\tilde{a}_3|\tilde{x}_3, \lambda), \\ P'(a_2|x_2, \lambda) &= P_{2\leftarrow 3}(a_2|x_2\tilde{x}_3\tilde{a}_3, \lambda). \end{aligned} \tag{5.10}$$

This argument holds equivalently for the l tripartite boxes and for any observer or input/output combination. Therefore, one can conclude that the tripartite boxes intervened in the preparation phase can be thought as shared randomness.

Let us know show that a similar statement can be made regarding the measurement phase. That is, that any sequence of wirings provides a final probability distribution with a local model along the bipartition $A_1 - A_2A_3$. For simplicity, we illustrate our procedure for the wiring shown in figure 5.3,

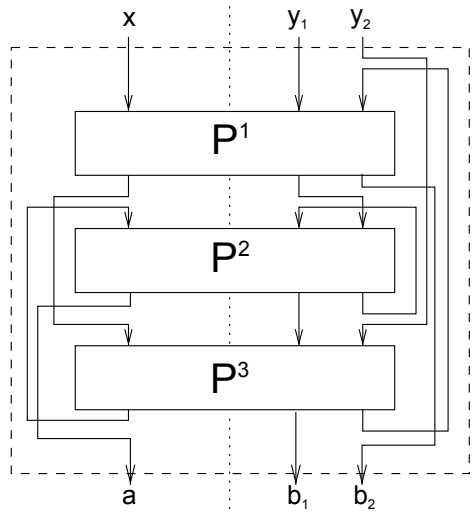

Figure 5.3: Wiring of several tripartite correlations distributed among parties A and B. The generated bipartite box accepts a bit x (two bits y_1, y_2) as input on subsystem A (B) and returns a bit a (two bits b_1, b_2) as output. Relations (5.11) guarantee that the final bipartite distribution $P_{\text{fin}}(a,(b_1,b_2)|x,(y_1,y_2))$ admits a local model.

where boxes P^1, P^2, P^3 are distributed between two parties A and B, and party A only holds one subsystem of each box. The construction is nevertheless general: it applies to any wiring and also covers situations where for some TOBL boxes party A holds two subsystems instead of just one (or even the whole box).

From (5.6) we have

$$\begin{aligned} P^i(a_1^i a_2^i a_3^i|x_1^i x_2^i x_3^i) &= \sum_{\lambda^i} p^i_{\lambda^i} P_1^i(a_1^i|x_1^i,\lambda^i) P^i_{2\to 3}(a_2^i a_3^i|x_2^i x_3^i,\lambda^i) \\ &= \sum_{\lambda^i} p^i_{\lambda^i} P_1^i(a_1^i|x_1^i,\lambda^i) P^i_{2\leftarrow 3}(a_2^i a_3^i|x_2^i x_3^i,\lambda^i), \end{aligned} \tag{5.11}$$

for $i = 1, 2, 3$. Consider the first box that receives an input, in our case subsystem 2 of P^1. The first outcome a_2^1 can be generated by the probability distribution $P^1_{2\to 3}(A_2^1, A_3^1|X_2^1, X_3^1, \lambda^1)$ encoded in the hidden variable λ^1 that models these first correlations. This is possible because for this decomposition a_2^1 is defined independently of x_3^1, the input in subsystem 3. Then, the next input x_3^2, which is equal to a_2^1, generates the output a_3^2 according to the probability distribution $P^2_{2\leftarrow 3}(A_2^2, A_3^2|X_2^2, X_3^2, \lambda^2)$ encoded in λ^2. The subsequent outcomes a_2^i and a_3^i are generated in a similar way. The general idea is that outputs are generated sequentially using the local models

according to the structure of the wiring on $2-3$. Finally, subsystem 1 can generate its outputs a^i by using the probability distribution $P_1^i(A_1^i|X^i,\lambda^i)$. This probability distribution is independent of the order in which parties 2 and 3 make their measurement choices for any of the boxes. Averaging over all hidden variables one obtains P_{fin}. This construction provides the desired local model for the final probability distribution. The analysis can be trivially extended to an arbitrary number of N-partite boxes. See appendix B.1 for a N-partite generalization of TOBL models. □

This theorem implies that TOBL models are consistent with the WCCPI formalism, and therefore it is to be considered as the adequate definition of multipartite nonlocality. On the other hand, the set of NSBL is as well consistent with the formalism, so one may ask whether TOBL set is indeed larger than the former. This is shown in the next theorem.

Theorem 5.2. *The set of TOBL correlations is strictly larger than the set of NSBL correlations.*

Proof. This theorem is shown by constructing an explicit example of a probability distribution which has a TOBL decomposition, however does not belong to the set of NSBL probability distributions. To prove this result, we consider the 'Guess Your Neighbor's Input' (GYNI) polynomial [ABB+10]

$$\begin{aligned}\beta_{\text{GYNI}}\cdot P(A_1,A_2,A_3|X_1,X_2,X_3) =& P(000|000)+P(110|011)\\ &+P(011|101)+P(101|110).\end{aligned} \tag{5.12}$$

The maximum of this quantity over the set of probabilities having a NSBL decomposition is equal to 1, that is $\beta_{\text{GYNI}}(P\in\text{NSBL})\leq 1$. In fact, consider the terms in the NSBL decomposition $P_1(a_1|x_1,\lambda)P_{23}^{\text{NS}}(a_2,a_3|x_2,x_3,\lambda)$. Without loss of optimality, one can restrict the analysis to correlations where $P_1(a_1|x_1,\lambda)$ is deterministic, say $P_1(0|0,\lambda)=P_1(0|1,\lambda)$. Thus, the GYNI polynomial for this set of probabilities satisfies

$$\begin{aligned}\beta_{\text{GYNI}}\cdot P(A_1,A_2,A_3|X_1,X_2,X_3,\lambda) &= P_{23}^{\text{NS}}(0,0|0,0,\lambda)+P_{23}^{\text{NS}}(1,1|0,1,\lambda)\\ &\leq P_2(0|0,\lambda)+P_2(1|0,\lambda)\leq 1\end{aligned} \tag{5.13}$$

with $P_2(a_2|x_2,\lambda)=\sum_{a_3}P_{23}^{\text{NS}}(a_2,a_3|x_2,x_3,\lambda)$ being a well-defined distribution due to the no-signaling constraints. One can easily check that the bound holds for any other deterministic choice of $P_1(0|0,\lambda)$ and $P_1(0|1,\lambda)$. As the NSBL decomposition is a convex mixture of these points, the GYNI polynomial is also bounded by 1. Note, however, that in Thm. 5.3 it is shown that there is a set of probabilities in TOBL obtaining larger values of the GYNI polynomial. Hence, the set of NSBL is strictly contained in TOBL. □

5.1.5 Conclusions

We have introduced a novel framework for the characterization of nonlocality which has an operational motivation and captures the role of nonlocality as a resource for device-independent quantum information processing. In spite of its simplicity, the framework questions the current understanding of genuine multipartite nonlocality, as the standard definition adopted by the community is inconsistent with it. Similar conclusions are reached from another perspective in [BPBG11]. We provide alternative frameworks where consistency is recovered. The main open question is now to identify the largest set of correlations that remain consistent under WCCPI protocols when some of the parties collaborate. We conjecture that TOBL correlations constitute such a set and, therefore, that for any bilocal model requiring two-way signaling terms there is a valid WCCPI protocol detecting its inconsistency.

5.2 Quantum correlations require multipartite information principles

An ubiquitous problem in Physics is to understand which correlations can be observed among different events. In fact, any theoretical model aims at predicting the experimental results of measurements, or actions, performed at different space-time locations. Naively, one could argue that any kind of correlations are in principle possible within a general physical theory, and that only the details of the devices used for establishing the correlations imply limitations on them. Interestingly, this intuition is not correct: general physical principles impose non-trivial constraints on the allowed correlations among distant observers, independently of any assumption on the internal working of the devices. It is then a crucial question to identify which correlations among distant observers are compatible with our current description of Nature based on Quantum Physics. In particular, it would be desirable to understand why some correlations cannot be realized by quantum means, even if they do not allow any faster-than-light communication [PR94].

Recently, information concepts have been advocated as the key missing ingredient needed to single-out the set of quantum correlations [Dam05, CBH03]. The main idea is to identify 'natural' information principles, formulated in terms only of correlations, which are satisfied by quantum correlations and proven to be violated by supra-quantum correlations. The existence of these supra-quantum correlations, then, would have implausible consequences from an information point of view. These information principles would provide a natural explanation of why the correlations observed in Nature have the quantum form. Celebrated examples of these principles are information causality [PPK$^+$09] or non-trivial communication complexity [Dam00, Dam05]. While the use of these information concepts has been successfully applied to some specific scenarios [BBL$^+$06, BS09, ABPS09, AKR$^+$10, CSS10], proving, or disproving, the validity of a principle for quantum correlations is extremely challenging. On the one hand, it is rather difficult to derive the Hilbert space structure needed for quantum correlations from information quantities. On the other hand, proving that some supra-quantum correlations are fully compatible with an information principle seems out of reach, as one needs to consider all possible protocols using these correlations and show that none of them leads to a violation of the principle. Thus, it is still open whether this approach is able to fully determine the set of quantum correlations.

In this section, we consider a general scenario consisting of an arbitrary

number of observers and show a fundamental limitation of this information-based program: no information principle based on bipartite concepts is able to determine the set of quantum correlations. Our results imply that determining the set of quantum correlations for an arbitrary number of observers, requires principles of an intrinsically multipartite structure.

5.2.1 Information principles

The analyzed scenario consists of n distant observers that can perform m possible measurements of d possible results on their systems. The observed correlations are described by the joint probability distribution $P(A_1, \ldots, A_n|X_1, \ldots, X_n)$, where $x_i = 0, \ldots, m-1$ denotes the measurement performed by party $i = 1, \ldots, n$; and $a_i = 0, \ldots, d-1$, the corresponding result. Each system is just seen as a black box producing the output a_i given the input x_i.

As explained in section 2.1.2, there exist three relevant sets of probability distributions: the no-signaling set $\mathcal{P}$, the local set $\mathcal{L}$, and the quantum set $\mathcal{Q}$. The first two sets have a clear interpretation. The set $\mathcal{P}$ represents the valid correlations under the principle of impossibility of faster-than-light communication. The set $\mathcal{L}$ represents the deterministic and local correlations.

On the other hand, the set $\mathcal{Q}$, despite having a clear mathematical definition (2.2), lacks a nice interpretation in terms of general principles, contrary to the classical and non-signaling counterparts. As said, it has been suggested that information concepts could provide the missing principles for quantum correlations.

It is worth mentioning before proceeding with the proof of the results that most of the existing examples of information principles have been formulated in the bipartite scenario. For example, information causality considers a scenario in which a first party, Alice, has a string of n_A bits. Alice is then allowed to send m classical bits to a second party, Bob. Information causality bounds the information Bob can gain on the n_A bits held by Alice whichever protocol they implement making use of the pre-established bipartite correlations and the message of m bits. Alice and Bob can violate this principle when they have access to some supra-quantum correlations [PPK$^+$09]. In the case $m = 0$, information causality implies that in absence of a message, pre-established correlations do not allow Bob to gain any information about any of the bits held by Alice, which is nothing but the no-signaling principle. The multipartite version of the no-signaling principle consists in the application of its bipartite version to all possible partitions

of the n parties into two groups. This suggests the following generalization of information causality to an arbitrary number of parties: given some correlations $P(a_1, \ldots, a_n|x_1, \ldots, x_n)$, they are said to be compatible with information causality whenever all bipartite correlations constructed from them satisfy this principle. This generalization ensures the correspondence between no-signaling and information causality when $m = 0$ for an arbitrary number of parties. This generalization of information causality has recently been applied to the study of extremal tripartite non-signaling correlations [YCA+12].

Regarding non trivial communication complexity, it studies how much communication is needed between two distant parties to compute probabilistically a function of some inputs in a distributed manner. It can also be interpreted as a generalization of the no-signaling principle, as it imposes constraints on correlations when a finite amount of communication is allowed between parties. Different multipartite generalizations of the principle have been studied, see [BvDHT99]. However, as for information causality, one can always consider the straightforward generalization in which the principle is applied to every partition of the n parties in two groups.

5.2.2 Supra-quantum correlations fulfilling information principles

In this section, we show that any physical principle that, similarly to no-signaling, is applied to every bipartition in the multipartite scenario is not sufficient to characterize the set of quantum correlations. We show this by finding tripartite correlations that, on one hand, fulfill any information principle based on bipartite concepts and, on the other hand, are supra-quantum.

Theorem 5.3. *There exist tripartite probability distributions that do not belong to the set of quantum correlations, however fulfill any information theoretic principle applied to all the bipartitions.*

Proof. This theorem is proven by finding an explicit example of a supra-quantum tripartite probability distribution that behaves locally under any bipartition. Any information principle has to be fulfilled by quantum correlations, and as a particular case, by local correlations. Therefore, the fact that the probability distribution is local in all bipartitions clearly implies that fulfills any information principle under any bipartition. Theorem 5.1 grants that a probability distribution with a TOBL decomposition with

a particular bipartition will behave locally under this bipartition. Furthermore, a tripartite probability distribution with a TOBL decomposition along any bipartitions will behave locally under any bipartition. In order to grant that the probability distribution does not have a quantum realization we will use the GYNI Bell inequality (5.12). Therefore, the whole problem of finding an example of a probability distribution proving Theorem 5.3 can reduced to the following numerical problem

$$\begin{aligned} \mathbf{B}_{\max} = \text{maximize } & \mathbf{B}(P) \\ \text{subject to} & \\ & P(A_1, A_2, A_3|X_1, X_2, X_3) \in \text{TOBL in all biparitions.} \end{aligned} \tag{5.14}$$

The maximization yields a value of $\mathbf{B}_{\max} = \frac{7}{6}$. As it has been previously mentioned, the maximum for quantum probability distributions is the unity. Therefore, the probability distribution maximizing the problem is supra-quantum and fulfills any information principle applied to the bipartition. Details of this probability distribution attaining the maximum of 7/6 and its TOBL decomposition can be found in Appendix B.2 . □

5.2.3 Conclusions

To summarize, we have shown that there exist tripartite non-signaling correlations that fulfill the principles of information causality and non-trivial communication complexity although they do not belong to the set of quantum correlations. The presented reasoning also applies to every other principle applied to the bipartitions of a multipartite system. This result provides a helpful insight for the formulation of a future principle aiming at distinguishing between quantum and supra-quantum correlations. In contrast to the no-signaling principle, such a forthcoming principle will need to be an intrinsically multipartite concept. This suggests that future research should be devoted to the development of information concepts of genuinely multipartite character. More specifically, one could investigate which multipartite generalizations of non trivial communication complexity can be considered intrinsically multipartite, and furthermore, how to generalize information causality for the case of multipartite communication protocols.

Chapter 6

Full randomness amplification

Understanding whether nature is deterministically pre-determined or there are intrinsically random processes is a fundamental question that has attracted the interest of multiple thinkers, ranging from philosophers and mathematicians to physicists or neuroscientists. Nowadays this question is also important from a practical perspective, as random bits constitute a valuable resource for applications such as cryptographic protocols, gambling, or the numerical simulation of physical and biological systems.

Classical physics is a deterministic theory. Perfect knowledge of the positions and velocities of a system of classical particles at a given time, as well as of their interactions, allows one to predict their future (and also past) behavior with total certainty [Lap40]. Thus, any randomness observed in classical systems is not intrinsic to the theory but just a manifestation of our imperfect description of the system.

The advent of quantum physics put into question this deterministic viewpoint, as there exist experimental situations for which quantum theory gives predictions only in probabilistic terms, even if one has a perfect description of the preparation and interactions of the system. A possible solution to this classically counterintuitive fact was proposed in the early days of quantum physics: Quantum mechanics had to be incomplete [EPR35], and there should be a complete theory capable of providing deterministic predictions for all conceivable experiments. There would thus be no room for intrinsic randomness, and any apparent randomness would again be a consequence of our lack of control over hypothetical "hidden variables" not contemplated by the quantum formalism.

Bell's no-go theorem [Bel64], however, implies that hidden-variable theories are inconsistent with quantum mechanics. Therefore, none of these could ever render a deterministic completion to the quantum formalism. More precisely, all hidden-variable theories compatible with a local causal structure predict that any correlations among space-like separated events satisfy a series of inequalities, known as Bell inequalities. Bell inequalities, in turn, are violated by some correlations among quantum particles. This form of correlations defines the phenomenon of quantum non-locality.

Now, it turns out that quantum non-locality does not necessarily imply the existence of fully unpredictable processes in nature. The reasons behind this are subtle. First of all, unpredictable processes could be certified only if the no-signaling principle holds. This states that no instantaneous communication is possible, which imposes in turn a local causal structure on events, as in Einstein's special relativity. In fact, Bohm's theory is both deterministic and able to reproduce all quantum predictions [Boh52], but it is incompatible with no-signaling. Thus, we assume throughout the validity of the no-signaling principle. Yet, even within the no-signaling framework, it is still not possible to infer the existence of fully random processes only from the mere observation of non-local correlations. This is due to the fact that Bell tests require measurement settings chosen at random, but the actual randomness in such choices can never be certified. The extremal example is given when the settings are determined in advance. Then, any Bell violation can easily be explained in terms of deterministic models. As a matter of fact, super-deterministic models, which postulate that all phenomena in the universe, including our own mental processes, are fully pre-programmed, are by definition impossible to rule out.

These considerations imply that the strongest result on the existence of randomness one can hope for using quantum non-locality is stated by the following possibility: Given a source that produces an arbitrarily small but non-zero amount of randomness, can one still certify the existence of completely random processes? Our main result is to provide an affirmative answer to this question. Our results, then, imply that the existence of correlations as those predicted by quantum physics forces us into a dichotomic choice: Either we postulate super-deterministic models in which all events in nature are fully pre-determined, or we accept the existence of fully unpredictable events.

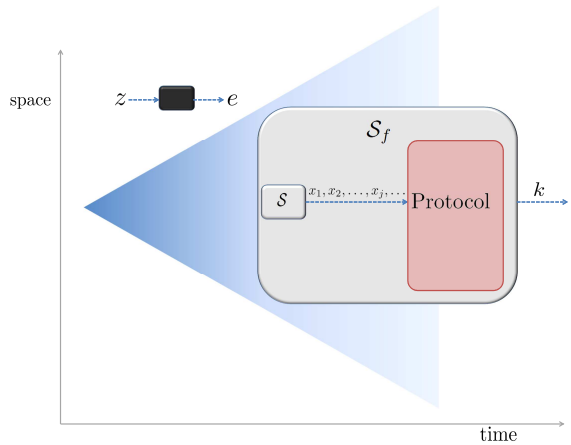

Figure 6.1: **Local causal structure and randomness amplification**. A source $\mathcal{S}$ produces a sequence $x_1, x_2, \ldots x_j, \ldots$ of imperfect random bits. The goal of randomness amplification is to produce a new source $\mathcal{S}_f$ of perfect random bits, that is, to process the initial bits to get a final bit k fully uncorrelated (free) from any potential cause of it. All space-time events outside the future light-cone of k may have been in its past light-cone before and therefore constitute a potential cause of it. Any such event can be modeled by a measurement z, with an outcome e, on some physical system. This system may be under the control of an adversary Eve, interested in predicting the value of k.

6.1 Randomness from an information science perspective

Besides the philosophical and physics-foundational implications, our results provide a protocol for perfect randomness amplification using quantum nonlocality. Randomness amplification is an information-theoretic task whose goal is to use an input source $\mathcal{S}$ of imperfectly random bits to produce perfect random bits that are arbitrarily uncorrelated from all the events that may have been a potential cause of them, i.e. arbitrarily free. In general, $\mathcal{S}$ produces a sequence of bits $x_1, x_2, \ldots x_j, \ldots$, with $x_j = 0$ or 1 for all j, see Fig. 6.1. Each bit j contains some randomness, in the sense that the probability $P(x_j|e)$ that it takes a given value x_j, conditioned on any pre-existing variable e, is such that

$$\epsilon \leq P(x_j|e) \leq 1 - \epsilon \tag{6.1}$$

for all j and e, where $0 < \epsilon \leq 1/2$. The variable e can correspond to any event that could be a possible cause of bit x_j. Therefore, e represents events

contained in the space-time region lying outside the future light-cone of x_j. Free random bits correspond to $\epsilon = \frac{1}{2}$; while deterministic ones, i.e. those predictable with certainty by an observer with access to e, to $\epsilon = 0$. More precisely, when $\epsilon = 0$ the bound (6.6) is trivial and no randomness can be certified. We refer to $\mathcal{S}$ as an ϵ-source, and to any bit satisfying (6.6) as an ϵ-free bit. The aim is then to generate, from arbitrarily many uses of $\mathcal{S}$, a final source $\mathcal{S}_f$ of ϵ_f arbitrarily close to 1/2. If this is possible, no cause e can be assigned to the bits produced by $\mathcal{S}_f$, which are then fully unpredictable. Note that efficiency issues, such as the rate of uses of $\mathcal{S}$ required per final bit generated by $\mathcal{S}_f$ do not play any role in randomness amplification. The relevant figure of merit is just the quality, measured by ϵ_f, of the final bits. Thus, without loss of generality, we restrict our analysis to the problem of generating a single final free random bit k.

Santha and Vazirani proved that randomness amplification is impossible using classical resources [SV86]. This is in a sense intuitive, in view of the absence of any intrinsic randomness in classical physics. In the quantum regime, randomness amplification has been recently studied by Colbeck and Renner [CR12]. There, $\mathcal{S}$ is used to choose the measurement settings by two distant observers, Alice and Bob, in a Bell test [BC90] involving two entangled quantum particles. The measurement outcome obtained by one of the observers, say Alice, in one of the experimental runs (also chosen with $\mathcal{S}$) defines the output random bit. Colbeck and Renner proved how input bits with very high randomness, of $0.442 < \epsilon \leq 0.5$, can be mapped into arbitrarily free random bits of $\epsilon_f \to 1/2$, and conjectured that randomness amplification should be possible for any initial randomness [CR12]. Our results also solve this conjecture, as we show that quantum non-locality can be exploited to attain *full randomness amplification*, i.e. that ϵ_f can be made arbitrarily close to 1/2 for any $0 < \epsilon \leq 1/2$.

Before presenting the ingredients of our proof, it is worth commenting on previous works on randomness in connection with quantum non-locality. In [PAM+10] it was shown how to bound the intrinsic randomness generated in a Bell test. These bounds can be used for device-independent randomness expansion, following a proposal by Colbeck [Col07], and to achieve a quadratic expansion of the amount of random bits (see [AMP12, PM11, FGS11, VV12] for further works on device-independent randomness expansion). Note however that, in randomness expansion, one assumes instead, from the very beginning, the existence of an input seed of free random bits, and the main goal is to expand this into a larger sequence. The figure of merit there is the ratio between the length of the final and initial strings of free random bits. Finally, other recent works have analyzed how a lack of random-

ness in the measurement choices affects a Bell test [KPB06, BG10, Hal10] and the randomness generated in it [KHS+12].

6.2 A protocol for full randomness amplification

Let us now sketch the realization of our final source $\mathcal{S}_f$. We use the input ϵ-source $\mathcal{S}$ to choose the measurement settings in a multipartite Bell test involving a number of observers that depends both on the input ϵ and the target ϵ_f. After verifying that the expected Bell violation is obtained, the measurement outcomes are combined to define the final bit k. For pedagogical reasons, we adopt a cryptographic perspective and assume the worst-case scenario where all the devices we use may have been prepared by an adversary Eve equipped with arbitrary non-signaling resources, possibly even supra-quantum ones. In the preparation, Eve may have also had access to $\mathcal{S}$ and correlated the bits it produces with some physical system at her disposal, represented by a black box in Fig. 6.1. Without loss of generality, we can assume that Eve can reveal the value of e at any stage of the protocol by measuring this system. Full randomness amplification is then equivalent to proving that Eve's correlations with k can be made arbitrarily small. In order to give a more intuitive description of our results, we first detail how a less strong result can be easily stated from basic properties of Bell inequalities.

6.2.1 Partial randomness from GHZ paradoxes

Bell tests for which quantum correlations achieve the maximal non-signaling violation, also known as Greenberger-Horne-Zeilinger (GHZ) paradoxes [GHZ89], are necessary for randomness amplification. This is due to the fact that unless the maximal non-signaling violation is attained, for sufficiently small ϵ, Eve may fake the observed correlations with classical deterministic resources.

Consider any correlations attaining the maximal violation of the five-party Mermin inequality [Mer90a]. In each run of this Bell test, measurements (inputs) $\mathbf{x} = (x_1, \ldots, x_5)$ on five distant black boxes generate 5 outcomes (outputs) $\mathbf{a} = (a_1, \ldots, a_5)$, distributed according to a non-signaling conditional probability distribution $P(\mathbf{a}|\mathbf{x})$. Both inputs and outputs are bits, as they can take two possible values, $x_i, a_i \in \{0, 1\}$ with $i = 1, \ldots, 5$. The inequality can be written as

$$\sum_{\mathbf{a},\mathbf{x}} I(\mathbf{a}, \mathbf{x}) P(\mathbf{a}|\mathbf{x}) \geq 6 \ , \tag{6.2}$$

with coefficients

$$I(\mathbf{a},\mathbf{x}) = (a_1 \oplus a_2 \oplus a_3 \oplus a_4 \oplus a_5)\,\delta_{\mathbf{x}\in\mathcal{X}_0} + (a_1 \oplus a_2 \oplus a_3 \oplus a_4 \oplus a_5 \oplus 1)\,\delta_{\mathbf{x}\in\mathcal{X}_1}\ , \quad (6.3)$$

where

$$\delta_{\mathbf{x}\in\mathcal{X}_0} = \begin{cases} 1 & \text{if } \mathbf{x}\in\mathcal{X}_0 \\ 0 & \text{if } \mathbf{x}\notin\mathcal{X}_0 \end{cases}\ ,$$

and

$$\begin{aligned}\mathcal{X}_0 &= \{(10000),(01000),(00100),(00010),(00001),(11111)\},\\ \mathcal{X}_1 &= \{(00111),(01011),(01101),(01110),(10011),(10101),\\ &\qquad (10110),(11001),(11010),(11100)\}.\end{aligned}$$

That is, only half of all possible combinations of inputs, namely those in $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$, appear in the Bell inequality.

The maximal, non-signaling and algebraic, violation of the inequality corresponds to the situation in which the left-hand side of (6.2) is zero. The key property of inequality (6.2) is that its maximal violation can be attained by quantum correlations. In fact, Mermin inequalities are defined for an arbitrary number of parties and quantum correlations attain the maximal non-signaling violation for any odd number of parties [Kly93]. This violation is always attained by performing local measurements on a GHZ quantum state.

Our interest in Mermin inequalities comes from the fact that, for an odd number of parties, they can be maximally violated by quantum correlations. These correlations, then, define a GHZ paradox, which is necessary for full randomness amplification. Nevertheless, GHZ paradoxes are however not sufficient. In fact, it is always possible to find non-signaling correlations that (i) maximally violate the 3-party Mermin inequality but (ii) assign a deterministic value to any function of the measurement outcomes. This observation can be checked for all unbiased functions mapping $\{0,1\}^3$ to $\{0,1\}$ (there are $\binom{8}{4}$ of those) through a linear program analogous to the one used to prove the next Theorem. For a larger number of parties, however, some functions cannot be deterministically fixed to an specific value while maximally violating a Mermin inequality, as implied by the following Theorem.

Theorem 6.1. *Let a five-party non-signaling conditional probability distribution $P(\mathbf{a}|\mathbf{x})$ in which inputs $\mathbf{x} = (x_1,\ldots,x_5)$ and outputs $\mathbf{a} = (a_1,\ldots,a_5)$ are bits. Consider the bit $\mathrm{maj}(\mathbf{a}) \in \{0,1\}$ defined by the majority-vote function of any subset consisting of three of the five measurement outcomes, say the first three, a_1, a_2 and a_3. Then, all non-signaling correlations attaining*

the maximal violation of the 5-party Mermin inequality are such that the probability that $\mathrm{maj}(\mathbf{a})$ *takes a given value, say 0, is bounded by*

$$1/4 \leq P\left(\mathrm{maj}(\mathbf{a}) = 0\right) \leq 3/4. \tag{6.4}$$

Proof. This result was obtained by solving a linear program. Therefore, the proof is numeric, but exact. Formally, let $P(\mathbf{a}|\mathbf{x})$ be a 5-partite non-signaling probability distribution. For $\mathbf{x} = \mathbf{x}_0 \in \mathcal{X}$, we performed the maximization,

$$\begin{aligned} P_{max} = \max_{P} \; & P(\mathrm{maj}(\mathbf{a}) = 0|\mathbf{x}_0) \\ & \text{subject to} \\ & I(\mathbf{a}, \mathbf{x}) \cdot P(\mathbf{a}|\mathbf{x}) = 0 \end{aligned} \tag{6.5}$$

which yields the value $P_{max} = 3/4$. Since the same result holds for $P(\mathrm{maj}(\mathbf{a}) = 1|\mathbf{x}_0)$, we get the bound $1/4 \leq P(\mathrm{maj}(\mathbf{a}) = 0) \leq 3/4$.

As a further remark, note that a lower bound to P_{max} can easily be obtained by noticing that one can construct conditional probability distributions $P(\mathbf{a}|\mathbf{x})$ that maximally violate 5-partite Mermin inequality (6.2) for which at most one of the output bits (say a_1) is deterministically fixed to either 0 or 1. If the other two output bits (a_2, a_3) were to be completely random, the majority-vote of the three of them $\mathrm{maj}(a_1, a_2, a_3)$ could be guessed with a probability of 3/4. Our numerical results say that this turns out to be an optimal strategy.

□

The partial unpredictability in the five-party Mermin Bell test is the building block of our protocol. To complete it, we must equip it with two essential components: (*i*) an *estimation procedure* that verifies that the untrusted devices do yield the required Bell violation; and (*ii*) a *distillation procedure* that, from sufficiently many ϵ_i-bits generated in the 5-party Bell experiment, distills a single final ϵ_f-source of $\epsilon_f \to 1/2$. To these ends, we consider a more complex Bell test involving N groups of five observers (quintuplets) each, as depicted in Fig. 6.2. The protocol is described in the next section.

In the Appendix 6 we prove using techniques from [Mas09] that, if the protocol is not aborted, the final bit produced by the protocol is indistinguishable from an ideal random bit uncorrelated to the eavesdropper. Thus, the output free random bits satisfy universally-composable security [Can01], the highest standard of cryptographic security, and could be used as seed for randomness expansion or any other protocol.

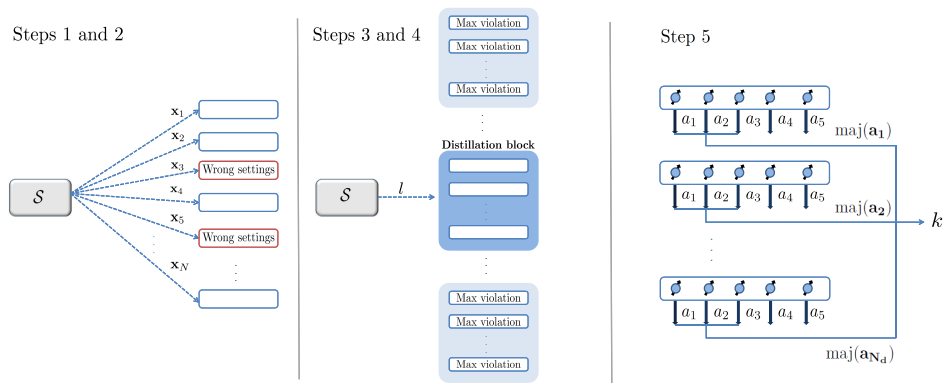

Figure 6.2: **Protocol for full randomness amplification based on quantum non-locality**. In the first two steps, all N quintuplets measure their devices, where the choice of measurement is done using the ϵ-source $\mathcal{S}$; the quintuplets whose settings happen not to take place in the five-party Mermin inequality are discarded (in red). In steps 3 and 4, the remaining quintuplets are grouped into blocks. One of the blocks is chosen as the distillation block, using again $\mathcal{S}$, while the others are used to check the Bell violation. In the fifth step, the random bit k is extracted from the distillation block.

6.2.2 Protocol for full randomness amplification

In this section, we describe the protocol to obtaine a fully random bit. The protocol uses as resources the ϵ-source $\mathcal{S}$ and $5N$ quantum systems. Recall that the bits produced by the source $\mathcal{S}$ are such that the probability $P(x_j|e)$ that bit j takes a given value x_j, conditioned on any pre-existing variable e, is bounded by

$$\epsilon \leq P(x_j|e) \leq 1-\epsilon, \tag{6.6}$$

for all j and e, where $0 < \epsilon \leq 1/2$. The bound, when applied to n-bit strings produced by the ϵ-source, implies that

$$\epsilon^n \leq P(x_1, \ldots, x_n|e) \leq (1-\epsilon)^n. \tag{6.7}$$

Each of the quantum systems is abstractly modeled by a black box with binary input x and output a. The protocol processes classically the bits generated by $\mathcal{S}$ and by the quantum boxes. The result of the protocol is a classical symbol k, associated to an abort/no-abort decision. If the protocol is not aborted, k encodes the final output bit, with possible values 0 or 1. Whereas when the protocol is aborted, no numerical value is assigned to k but the symbol $\varnothing$ instead, representing the fact that the bit is empty. The formal steps of the protocol are:

1. $\mathcal{S}$ is used to generate N quintuple-bits $\mathbf{x}_1, \dots \mathbf{x}_N$, which constitute the inputs for the $5N$ boxes. The boxes then provide N output quintuple-bits $\mathbf{a}_1, \dots \mathbf{a}_N$.

2. The quintuplets such that $\mathbf{x} \notin \mathcal{X}$ are discarded. The protocol is aborted if the number of remaining quintuplets is less than $N/3$.

3. The quintuplets left after step 2 are organized in N_b blocks each one having N_d quintuplets. The number N_b of blocks is chosen to be a power of 2. For the sake of simplicity, we relabel the index running over the remaining quintuplets, namely $\mathbf{x}_1, \dots \mathbf{x}_{N_b N_d}$ and outputs $\mathbf{a}_1, \dots \mathbf{a}_{N_b N_d}$. The input and output of the j-th block are defined as $y_j = (\mathbf{x}_{(j-1)N_d+1}, \dots \mathbf{x}_{(j-1)N_d+N_d})$ and $b_j = (\mathbf{a}_{(j-1)N_d+1}, \dots \mathbf{a}_{(j-1)N_d+N_d})$ respectively, with $j \in \{1, \dots, N_b\}$. The random variable $l \in \{1, \dots N_b\}$ is generated by using $\log_2 N_b$ further bits from $\mathcal{S}$. The value of l specifies which block (b_l, y_l) is chosen to generate k, i.e. the distilling block. We define $(\tilde{b}, \tilde{y}) = (b_l, y_l)$. The other $N_b - 1$ blocks are used to check the Bell violation.

4. The function

$$r[b, y] = \begin{cases} 1 & \text{if } I(\mathbf{a}_1, \mathbf{x}_1) = \cdots = I(\mathbf{a}_{N_d}, \mathbf{x}_{N_d}) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.8}$$

tells whether block (b, y) features the right correlations ($r = 1$) or the wrong ones ($r = 0$), in the sense of being compatible with the maximal violation of inequality (6.2). This function is computed for all blocks but the distilling one. The protocols is aborted unless all of them give the right correlations,

$$g = \prod_{j=1, j \neq l}^{N_b} r[b_j, y_j] = \begin{cases} 1 & \text{not abort} \\ 0 & \text{abort} \end{cases} . \tag{6.9}$$

Note that the abort/no-abort decision is independent of whether the distilling block l is right or wrong.

5. If the protocol is not aborted then k is assigned a bit generated from $b_l = (\mathbf{a}_1, \dots \mathbf{a}_{N_d})$ as

$$k = f(\mathrm{maj}(\mathbf{a}_1), \dots \mathrm{maj}(\mathbf{a}_{N_d})) \ . \tag{6.10}$$

Here $f : \{0,1\}^{N_d} \to \{0,1\}$ is a function characterized in Lemma C.4 below, while $\mathrm{maj}(\mathbf{a}_i) \in \{0,1\}$ is the majority-vote among the three first bits of the quintuple string $\mathbf{a}_i$. If the protocol is aborted it sets $k = \varnothing$.

At the end of the protocol, k is potentially correlated with the settings of the distilling block $\tilde{y} = y_l$, the bit g in (6.9), and the bits

$$t = [l, (b_1, y_1), \dots (b_{l-1}, y_{l-1}), (b_{l+1}, y_{l+1}), \dots (b_{N_b}, y_{N_b})].$$

Additionally, an eavesdropper Eve might have a physical system correlated with k, which she may measure at any instance of the protocol. This system is not necessarily classical or quantum, the only assumption about it is that measuring it does not produce instantaneous signaling anywhere else. We label all possible measurements Eve can perform with the classical variable z, and with e the corresponding outcome. In summary, after the performance of the protocol all the relevant information is $k, \tilde{y}, t, g, e, z$, with statistics described by an unknown conditional probability distribution $P(k, \tilde{y}, t, g, e|z)$.

To assess the security of our protocol for full randomness amplification, we have to show that the distribution describing the protocol when not aborted is indistinguishable from the distribution $P_{\mathrm{ideal}}(k, \tilde{y}, t, g, e|zg = 1) = \frac{1}{2}P(\tilde{y}, t, e|zg = 1)$ describing an ideal free random bit. For later purposes, it is convenient to cover the case when the protocol is aborted with an equivalent notation: if the protocol is aborted, we define $P(k, \tilde{y}, t, e|zg = 0) = \delta_k^{\varnothing} P(\tilde{y}, t, e|zg = 0)$ and $P_{\mathrm{ideal}}(k, \tilde{y}, t, e|zg = 0) = \delta_k^{\varnothing} P(\tilde{y}, t, e|zg = 0)$, where $\delta_k^{k'}$ is a Kronecker's delta. In this case, it is immediate that $P = P_{\mathrm{ideal}}$, as the locally generated symbol $\varnothing$ is always uncorrelated to the environment. To quantify the indistinguishability between P and P_{ideal}, we consider the scenario in which an observer, having access to all the information $k, \tilde{y}, t, g, e, z$, has to correctly distinguish between these two distributions. We denote by $P(\mathrm{guess})$ the optimal probability of correctly guessing between the two dis-

tributions. This probability reads

$$P(\text{guess}) = \frac{1}{2} + \frac{1}{4} \sum_{k,\tilde{y},t,g} \max_z \sum_e \Big| P(k,\tilde{y},t,g,e|z) - P_{\text{ideal}}(k,\tilde{y},t,g,e|z) \Big|, \tag{6.11}$$

where the second term can be understood as (one fourth of) the variational distance between P and P_{ideal} generalized to the case when the distributions are conditioned on an input z [Mas09]. If the protocol is such that this guessing probability can be made arbitrarily close to 1/2, it generates a distribution P that is basically undistinguishable from the ideal one. This is known as "universally-composable security", and accounts for the strongest notion of cryptographic security (see [Can01] and [Mas09]). It implies that the protocol produces a random bit that is secure (free) in any context. In particular, it remains secure even if the adversary Eve has access to $\tilde{y}$, t and g.

Our main result, namely the security of our protocol for full randomness amplification, follows from the following Theorem.

Theorem 6.2 (**Main Theorem**). *Consider the previous protocol for randomness amplification and the conditional probability distribution $P(k,\tilde{y},t,g,e|z)$ describing the statistics of the bits $k,\tilde{y},t,g$ generated during its execution and any possible system with input z and output e correlated to them. The probability $P(\text{guess})$ of correctly guessing between this distribution and the ideal distribution $P_{\text{ideal}}(k,\tilde{y},t,g,e|z)$ is such that*

$$P(\text{guess}) \;\leq\; \frac{1}{2} + \frac{3\sqrt{N_d}}{2} \left[\alpha^{N_d} + 2\, N_b^{\log_2(1-\epsilon)} \left(32\beta\epsilon^{-5}\right)^{N_d} \right] \;. \tag{6.12}$$

where α and β are real numbers such that $0 < \alpha < 1 < \beta$.

The right-hand side of (6.12) can be made arbitrary close to 1/2, for instance by setting $N_b = \left(32\,\beta\,\epsilon^{-5}\right)^{2N_d/|\log_2(1-\epsilon)|}$ and increasing N_d subject to the fulfillment of the condition $N_d N_b \geq N/3$. [Note that $\log_2(1-\epsilon) < 0$.] In the limit $P(\text{guess}) \to 1/2$, the bit k generated by the protocol is indistinguishable from an ideal free random bit.

The proof of Theorem 6.2 is provided in appendix C. Here, we comment on the main intuitions behind our protocol. As mentioned, the protocol builds on the 5-party Mermin inequality because it is the simplest GHZ paradox allowing some randomness certification. The estimation part, given by step 4, is rather standard and inspired by estimation techniques introduced in [BKP06], which were also used in [CR12] in the context of randomness amplification. The most subtle part is the distillation of the final

bit in step 5. Naively, and leaving aside estimation issues, one could argue that it is nothing but a classical processing by means of the function f of the imperfect random bits obtained via the N_d quintuplets. But this seems in contradiction with the result by Santha and Vazirani proving that it is impossible to extract by classical means a perfect free random bit from imperfect ones [SV86]. This intuition is however wrong. The reason is because in our protocol the randomness of the imperfect bits is certified by a Bell violation, which is impossible classically. Indeed, the Bell certification allows applying techniques similar to those obtained in Ref. [Mas09] in the context of privacy amplification against non-signaling eavesdroppers. There, it was shown how to amplify the privacy, that is the unpredictability, of one of the measurement outcomes of bipartite correlations violating a Bell inequality. The key point is that the amplification, or distillation, was attained in a *deterministic* manner. That is, contrary to standard approaches, the privacy amplification process described in [Mas09] does not consume any randomness. Clearly, these deterministic techniques are extremely convenient for our randomness amplification scenario. In fact, the distillation part in our protocol can be seen as the translation of the privacy amplification techniques of Ref. [Mas09] to our more complex scenario, involving now 5-party non-local correlations and a function of three of the measurement outcomes.

To end up with, we must show that quantum resources can indeed successfully implement our protocol. It is immediate to see that the qubit measurements X or Y on the quantum state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00000\rangle + |11111\rangle)$, with $|0\rangle$ and $|1\rangle$ the eigenstates of the Z qubit basis, yield correlations that maximally violate the five-partite Mermin inequality in question. This completes our main result.

6.3 Conclusions

In summary, we have presented a protocol that, using quantum non-local resources, attains *full randomness amplification*. This task is impossible classically and was not known to be possible in the quantum regime. As our goal was to prove full randomness amplification, our analysis focuses on the noise-free case. In fact, the noisy case only makes sense if one does not aim at perfect random bits and bounds the amount of randomness in the final bit. Then, it should be possible to adapt our protocol in order to get a bound on the noise it tolerates. Other open questions that naturally follow from our results consist of studying randomness amplification against quantum eavesdroppers, or the search of protocols in the bipartite scenario.

From a more fundamental perspective, our results imply that there exist experiments whose outcomes are fully unpredictable. The only two assumptions for this conclusion are the existence of events with an arbitrarily small but non-zero amount of randomness and the validity of the no-signaling principle. Dropping the former implies accepting a super-determinisitc view where no randomness exist, so that we experience a fully pre-determined reality. This possibility is uninteresting from a scientific perspective, and even uncomfortable from a philosophical one. Dropping the latter, in turn, implies abandoning a local causal structure for events in space-time. However, this is one of the most fundamental notions of special relativity, and without which even the very meaning of randomness or predictability would be unclear, as these concepts implicitly rely on the cause-effect principle.

Chapter 7

Conclusions and outlook

In this thesis we have introduced novel applications of the device-independent formalism and achieved significant improvements in other applications that were previously known. Also, we have studied the phenomenon of nonlocality, focusing on quantum correlations that are maximally nonlocal and multipartite nonlocality. In this section we review on further research and open venues that the work in this thesis leads to.

Dimensionality

Our contribution was to develop tests to estimate the dimensionality of an unknown physical system. In analogy with Bell inequalities, we built a family of linear inequalities that allow one to lower bound the dimensionality required to describe the experiment. We studied the set of valid probability distributions for classical and quantum systems of a given dimension. This was done under the assumption that the source and the measurement shared classical correlations or no correlations at all. A natural extension to be further studied is a scenario in which the devices are allowed to share quantum correlations. Clearly, this would provide a enlarged set of probability distributions. A source sending classical systems but allowing for quantum correlations would coincide with the scenario of entanglement assisted random access codes (EARAC) [PZ10]. It is interesting to study how results in EARAC can be translated into classical dimension witnesses in the presence of quantum correlations. A more general scenario is to study sources sending quantum systems in the presence of quantum correlations. This is related with the scenario considered in super-dense coding [BW92]. Can one estimate the dimensionality of the pre-established correlations and the message sent? How does the gap between classical and quantum messages

behave in a scenario where quantum correlations are allowed?

Maximally nonlocal correlations

We have studied quantum systems providing correlations that are maximally nonlocal. First, we have studied the bipartite scenario. Our main contribution was to provide a recipe to design maximally nonlocal experiments. We have conducted an experiment on entangled photons to obtain the most nonlocal correlations ever reported. An immediate challenge remains to perform an experiment providing almost fully nonlocal correlations. More interestingly, one may study which properties of the correlations obtained by our recipe are useful for information tasks such as cryptography or randomness expansion. Some of the examples we provided have been studied in this direction, however providing negative results. One may consider how these useful properties depend on the Kochen-Specker set used as a primitive and perhaps engineer new sets that are useful.

In the multipartite setting, our main contribution is to show that quantum mechanics provide correlations that are maximally genuine nonlocal, monogamous and locally random. These properties make them appealing for multipartite information protocols. Indeed, we have shown that are useful to implement device-independent secret-sharing. However, the Bell test we have designed requires an absurdly large number of measurements performed on the systems and it is not robust against noise. Interestingly, this drawback was also present in the first device-independent protocols reported on cryptography [BHK05]. It is of a clear interest to design similar multipartite nonlocal correlations in a more experimental-friendly manner.

Operational framework for nonlocality and information-theoretic principles

Our contribution was to study nonlocality in the framework of resource theories. We have shown that nonlocality can be defined as a property that is preserved against a set of operations that includes wirings and prior to inputs classical communication. In the multipartite setting, we have shown that the current definition of multipartite nonlocality is inconsistent with the operational framework. We have proposed new definitions to recover consistency. It remains open whether one can find new definitions of multipartite nonlocality that are consistent with the formalism but that impose less constraints than TOBL models. On the other hand, our formalism allows one to reveal multipartite nonlocality of new probability distributions

and possibly, new quantum states. Our formalism may be useful to establish a sharpened link between multipartite entanglement and multipartite nonlocality.

Regarding information-theoretic principles, our main contribution was to show that the most promising candidates are insufficient to bound quantum correlations. In particular, we showed that information-theoretic principles cannot be successful if applied to the bipartitions of a multipartite system. It remains open to find an operational principle that takes into account these limitations.

Full randomness amplification

We have addressed the problem of certifying a full random process. Quantum theory makes probabilistic predictions, however a forthcoming deterministic theory may complete quantum predictions. Thus, randomness has to be certified independent of the theoretical framework. This question is tackled by Bell's theorem, however it requires an initial source of perfect randomness. Our contribution is to remove the assumption on the initial source of random bits. Our protocol only requires initial bits providing an arbitrarily small, but nonzero, amount of initial randomness. The implications for foundation of physics are extremely profound: our world is either completely deterministic or there exist fully unpredictable processes. Our result is to be understood as a thought experiment, therefore it is shown to work in an idealized noise-free experiment. A natural extension would be to provide similar protocols that are efficient in realistic situations.

Final remarks

The vast majority of device-independent protocols are based in nonlocality. However, in spite of large improvement of experimental capabilities, a definitive experiment showing nonlocality is still missing. This is mainly due to the imperfection of the devices and the detection loophole. There exist two alternatives for the scientific community: (i) we continue developing device-independent protocols relying on nonlocality and hope for a forthcoming resolution of the detection loophole (ii) we develop not-so-device-independent protocols that overcome the detection loophole, and that require a less demanding experimental implementation in comparison with nonlocality. Alternative (i) has been vastly followed by the community and this thesis is a good example. Nevertheless, an approach following (ii) is stated in section 3.2. Although it is a very specific result concerning dimension witness, we

believe that this kind of approaches, where extra reasonable assumptions are imposed, are the future of device-independent protocols.

Appendix A

Proofs of section 4.2

A.1 Proof of Theorem 4.8

To end up with, we provide here just the main steps of the proof, the most technical calculations being detailed in the appendices.

The proof for arbitrary $N \geq 2$ is in a similar spirit to the proof given in [BKP06] for the particular case $N = 2$. We proceed by *reductio ad absurdum*. We start by the particular marginal $P(r_A, ..., r_Y | \alpha, \alpha + \beta - 1, ..., \chi + \psi - 1)$, corresponding to all parts but Z, and assume that for some input $(\alpha', \alpha' + \beta' - 1, ..., \chi' + \psi' - 1)$ the most probable outcome $(a^{max}_{(\alpha',...,\psi')}, ..., y^{max}_{(\alpha',...,\psi')})$ is such that

$$P(a^{max}_{(\alpha',...,\psi')}, ..., y^{max}_{(\alpha',...,\psi')} | \alpha', \alpha'+\beta'-1, ..., \chi'+\psi'-1) > 1/d^{N-1} + \frac{d(N-1)}{4}\varepsilon. \tag{A.1}$$

Then, we prove that this implies that $\overline{I^N_M}(P) > \varepsilon$, which contradicts the hypothesis. [The same assumption for $(\alpha' + 1, \alpha' + \beta' - 1, ..., \chi' + \psi' - 1)$ would lead to an equivalent contradiction.] Finally, we extend the proof to the other $(N-1)$-partite marginals by symmetry.

First, since

$$\langle [W] \rangle = \sum_{i=1}^{d-1} iP\big([W] = i\big) \geq 1 - P\big([W] = 0\big), \tag{A.2}$$

we see from (4.19) that

$$\overline{I_M^N} \geq 2M - \frac{1}{M^{N-2}} \cdot$$
$$\cdot \sum_{\alpha,\beta,...,\chi,\psi=1}^{M} \Big(P\big(A_\alpha = [B_{\alpha+\beta-1} - ... + (-1)^{N-1} Y_{\chi+\psi-1} - (-1)^{N-1} Z_\psi]\big) +$$
$$P\big(A_{\alpha+1} = [B_{\alpha+\beta-1} - ... + (-1)^{N-1} Y_{\chi+\psi-1} - (-1)^{N-1} Z_\psi]\big)\Big). \tag{A.3}$$

Next, we notice in section A.4 that, for all $(\alpha, ..., \omega)$,

$$\begin{aligned} &P\big(A_\alpha = [B_\beta - ... + (-1)^{N-1} Y_\psi - (-1)^{N-1} Z_\zeta]\big) \leq \\ &1 - \big| P(A_\alpha = a, B_\beta = b, ..., Y_\psi = y) \\ &- P\big(B_\beta = b, ..., Y_\psi = y, Z_\zeta = [y - ... + (-1)^{N-1} b - (-1)^{N-1} a]\big)\big|, \end{aligned} \tag{A.4}$$

for any $(a, b, ..., y)$. Then,

$$\begin{aligned} \overline{I_M^N} \geq \frac{1}{M^{N-2}} \sum_{\alpha,\beta,...,\chi,\psi=1}^{M} &\big| P\big(A_\alpha = a, B_{\alpha+\beta-1} = b, ..., Y_{\chi+\psi-1} = y\big) \\ &- P\big(A_{\alpha+1} = a, B_{\alpha+\beta-1} = b, ..., Y_{\chi+\psi-1} = y\big)\big|, \end{aligned} \tag{A.5}$$

where the triangle inequality has been used.

In section A.5, in turn, we see that hypothesis (A.1) implies that there exists some point $(a_{0(\alpha',...,\psi')}, ..., y_{0(\alpha',...,\psi')})$ in the d^{N-1}-dimensional cubic grid $\mathcal{G} = \{0, 1, ..., d-1\}^{\times(N-1)}$, such that

$$\begin{aligned} &\big| P(A_{\alpha'} = a_{0(\alpha',...,\psi')}, ..., Y_{\chi'+\psi'-1} = y_{0(\alpha',...,\psi')}) \; - \\ &P(A_{\alpha'} = \dot{a}_{0(\alpha',...,\psi')}, ..., Y_{\chi'+\psi'-1} = \dot{y}_{0(\alpha',...,\psi')})\big| \; > \; \varepsilon, \end{aligned} \tag{A.6}$$

where $(\dot{a}_{0(\alpha',...,\psi')}, ..., \dot{y}_{0(\alpha',...,\psi')}) \in \mathcal{G}$ is any nearest neighbor of $(a_{0(\alpha',...,\psi')}, ..., y_{0(\alpha',...,\psi')})$.

Now, defining $\tilde{\beta} = \alpha + \beta$, ..., $\tilde{\chi} = \varphi + \chi$ and $\tilde{\psi} = \chi + \psi$ in (A.5), we see that

$$\begin{aligned} \overline{I_M^N} \geq \frac{1}{M^{N-2}} \sum_{\tilde{\psi},...,\tilde{\beta}=1}^{M} \Big| \sum_{\alpha=1}^{M} &\big(P(A_\alpha = a, B_{\tilde{\beta}-1} = b, ... Y_{\tilde{\psi}-1} = y) \\ &- P(A_{\alpha+1} = a, B_{\tilde{\beta}-1} = b, ..., Y_{\tilde{\psi}-1} = y)\big)\Big|. \end{aligned} \tag{A.7}$$

Here, we choose $b \equiv b_{0(\alpha',...,\psi')}$, ..., and $y \equiv y_{0(\alpha',...,\psi')}$. In turn, we set $a = a_{0(\alpha',...,\psi')}$, for all $1 \leq \alpha \leq \alpha'$, and $a = a_{0(\alpha',...,\psi')} + 1$, for all $\alpha' + 1 \leq \alpha \leq M$. With this, inequality (A.7) becomes

$$\begin{aligned}
\overline{I_M^N} &\geq \frac{1}{M^{N-2}} \sum_{\tilde{\psi},...,\tilde{\beta}=1}^{M} \big|(P(A_\alpha = a_{0(\alpha',...,\psi')}, B_{\tilde{\beta}-1} = b_{0(\alpha',...,\psi')}, ..., Y_{\tilde{\psi}-1} = y_{0(\alpha',...,\psi')}) \\
&- P(A_\alpha = a_{0(\alpha',...,\psi')} + 1, B_{\tilde{\beta}-1} = b_{0(\alpha',...,\psi')}, ..., Y_{\tilde{\psi}-1} = y_{0(\alpha',...,\psi')})\big| \\
&> \frac{1}{M^{N-2}} \sum_{\tilde{\psi},...,\tilde{\beta}=1}^{M} \varepsilon = \varepsilon,
\end{aligned}$$

where we have used that $A_{M+1} = [A_1 + 1]$ and invoked property (A.6). The last inequality finishes the proof for marginal $P(r_A, ..., r_Y|\alpha, \alpha+\beta-1, ..., \chi+\psi-1)$.

The proof for any other $(N-1)$-party marginal that includes A is a replica but where, before (A.5), instead of grouping it together with B, C, ..., and Y, one groups A with any choice of $N-2$ out of the other $N-1$. Finally, the proof for the marginal $P\big(r_B, ..., r_Z|\alpha, \alpha+\beta-1, ..., \chi+\psi-1\big)$, follows due to invariance of $\overline{I_M^N}$ under the exchange of A with, for instance, C (see section A.2). □

A.2 Symmetry under permutations of parts

Here we show the invariance of the N-partite inequality (4.19) under certain permutations of parts, the same arguments holding also for the tripartite case of (4.17). We show explicitly that (4.19) is symmetric with respect to the exchange of the N-th and the $(N-2)$-th parts, Z and X. The proof for the exchange $A \leftrightarrow C$ is exactly the same. We write the Bell polynomial of (4.19) as

$$\begin{aligned}
\overline{I_M^N} = \frac{1}{M^{N-2}} \sum_{\alpha,\beta,...,\varphi,\chi,\psi=1}^{M} &\big(J_{\alpha,...,\varphi,\chi,\psi}(A, B, ..., X, Y, Z) \\
&+ H_{\alpha,...,\varphi,\chi,\psi}(A, B, ..., X, Y, Z)\big),
\end{aligned} \tag{A.8}$$

with

$$\begin{aligned}
&J_{\alpha,...,\varphi,\chi,\psi}(A, B, ..., X, Y, Z) = \\
&\langle[A_\alpha - B_{\alpha+\beta-1} \; ... + (-1)^{N-1} X_{\varphi+\chi-1} - (-1)^{N-1} Y_{\chi+\psi-1} + (-1)^{N-1} Z_\psi]\rangle,
\end{aligned}$$

$$H_{\alpha,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z) =$$
$$\langle [B_{\alpha+\beta-1} - A_{\alpha+1} \; ... + (-1)^N X_{\varphi+\chi-1} - (-1)^N Y_{\chi+\psi-1} + (-1)^N Z_\psi] \rangle.$$

Under the exchange $X \leftrightarrow Z$, these matrices transform as

$$\begin{aligned} J_{\alpha,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z) \rightarrow & J_{\alpha,...,\varphi,\chi,\psi}(A,B,...,Z,Y,X) \\ & \equiv J_{\alpha,...,\varphi,\psi-\varphi+1,\varphi+\chi-1}(A,B,...,X,Y,Z) \end{aligned}$$

$$\begin{aligned} H_{\alpha,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z) \rightarrow & H_{\alpha,...,\varphi,\chi,\psi}(A,B,...,Z,Y,X) \\ & \equiv H_{\alpha,...,\varphi,\psi-\varphi+1,\varphi+\chi-1}(A,B,...,X,Y,Z). \end{aligned}$$

We notice that, due to the symmetries in the definition of matrix J, the fact that $\Omega_{i\times M+\omega} = [\Omega_\omega + i]$, for any $\Omega = A, B, C, ..., Y$, or Z, implies that

$$J_{\alpha,...,\omega\pm M,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z) = J_{\alpha,...,\omega,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z)$$

for any $\omega = \alpha, \beta, ...\varphi, \chi$ or ψ. Analogously, the same property holds also for matrix H. Hence, we have that

$$\begin{aligned} & \sum_{\alpha,\beta,...,\varphi,\chi,\psi=1}^{M} J_{\alpha,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z) \\ \equiv & \sum_{\alpha,\beta,...,\varphi,\chi,\psi=1}^{M} J_{\alpha,...,\varphi,\psi-\varphi+1,\varphi+\chi-1}(A,B,...,X,Y,Z) \end{aligned}$$

and

$$\begin{aligned} & \sum_{\alpha,\beta,...,\varphi,\chi,\psi=1}^{M} H_{\alpha,...,\varphi,\chi,\psi}(A,B,...,X,Y,Z) \\ \equiv & \sum_{\alpha,\beta,...,\varphi,\chi,\psi=1}^{M} H_{\alpha\;...,\varphi,\psi-\varphi+1,\varphi+\chi-1}(A,B,...,X,Y,Z). \end{aligned}$$

Therefore, it is $\overline{I_M^N} \equiv \overline{I_M^N}(X \leftrightarrow Z)$. $\square$

A.3 Quantum realization. Equation (4.21)

Here we prove that, for any $N > 2$, it is $\overline{I_M^N}(\Psi_d^N) = \overline{I_M^{N-1}}(\Psi_d^{N-1})$. Consider first the expectation value

$$\begin{aligned}
&\langle\Psi_d^{N-1}|[\hat{A}_\alpha - \hat{B}_{\alpha+\beta-1} + ... - (-1)^{N-1}\hat{Y}_\chi]|\Psi_d^{N-1}\rangle \\
&\equiv \sum_{r_{A_\alpha}, r_{B_{\alpha+\beta-1}}, ..., r_{Y_\chi}=0}^{d-1} [r_{A_\alpha} - r_{B_{\alpha+\beta-1}} + ... - (-1)^{N-1} r_{Y_\chi}] \big|\langle r_{A_\alpha}|\langle r_{B_{\alpha+\beta-1}}|...\langle r_{Y_\chi}|\Psi_d^{N-1}\rangle\big|^2 \\
&= \frac{1}{d^N} \times d^{N-2} \sum_{n=1}^{d-1} n \Big|\frac{1 - e^{-\pi i/M}}{1 - e^{-\frac{2\pi i}{d}(n+1/2)}}\Big|^2,
\end{aligned} \tag{A.9}$$

where we have used the explicit definitions of $|r_{A_\alpha}\rangle$, $|r_{B_{\alpha+\beta-1}}\rangle$, ..., and $|r_{Y_\chi}\rangle$, summed a geometric sequence, and introduced $n \equiv r_{A_\alpha} - r_{B_{\alpha+\beta-1}} + ... - (-1)^{N-1} r_{Y_\chi}$. Consider next

$$\begin{aligned}
&\langle\Psi_d^{N}|[\hat{A}_\alpha - \hat{B}_{\alpha+\beta-1} + ... + (-1)^{N-1}\hat{Z}_\psi]|\Psi_d^{N}\rangle \\
&\equiv \sum_{r_{A_\alpha}, r_{B_{\alpha+\beta-1}}, ..., r_{Z_\psi}=0}^{d-1} [r_{A_\alpha} - r_{B_{\alpha+\beta-1}} + ... + (-1)^{N-1} r_{Z_\psi}] \big|\langle r_{A_\alpha}|\langle r_{B_{\alpha+\beta-1}}|...\langle r_{Z_\psi}|\Psi_d^{N}\rangle\big|^2 \\
&= \frac{1}{d^{N+1}} \times d^{N-1} \sum_{n'=1}^{d-1} n' \Big|\frac{1 - e^{-\pi i/M}}{1 - e^{-\frac{2\pi i}{d}(n'+1/2)}}\Big|^2,
\end{aligned} \tag{A.10}$$

where we have now also used the definition of $|r_{Z_\psi}\rangle$, and introduced $n' \equiv r_{A_\alpha} - r_{B_{\alpha+\beta-1}} + ... + (-1)^{N-1} r_{Z_\psi}$. Notice that both expectation values coincide for any α, β, ..., χ, and ψ. In addition, the same analysis holds true for $\langle\Psi_d^{N-1}|[\hat{B}_{\alpha+\beta-1} - \hat{A}_{\alpha+1} - ... + (-1)^{N-1}\hat{Y}_\chi]|\Psi_d^{N-1}\rangle$ and $\langle\Psi_d^{N}|[\hat{B}_{\alpha+\beta-1} - \hat{A}_{\alpha+1} - ... - (-1)^{N-1}\hat{Z}_\psi]|\Psi_d^{N}\rangle$. Therefore, the Bell value of the ψ-th term of $\overline{I_M^N}$, $\frac{1}{M^{N-2}}\sum_{\alpha,\beta,...,\chi=1}^{M}\big(\langle[A_\alpha - B_{\alpha+\beta-1} + ... - (-1)^{N-1}Y_{\chi+\psi-1} + (-1)^{N-1}Z_\psi]\rangle + \langle[B_{\alpha+\beta-1} - A_{\alpha+1} - ... + (-1)^{N-1}Y_{\chi+\psi-1} - (-1)^{N-1}Z_\psi]\rangle\big)$, obtained from quantum measurements on $|\Psi_d^N\rangle$, is equal to $1/M$ times the Bell value of $\overline{I_M^{N-1}}$ obtained from $|\Psi_d^{N-1}\rangle$. Summing over ψ completes the proof. □

A.4 Proof of bound (A.4)

Here we show that, for all $(\alpha, \beta, ..., \psi, \zeta)$, and any $(a, b, ..., y)$, bound (A.4) holds. One has

$$
\begin{aligned}
& P\big(A_\alpha = B_\beta - ... + (-1)^{N-1} Y_\psi - (-1)^{N-1} Z_\zeta\big) \\
= & \sum_{i,j,...,m} P\big(A_\alpha = i, B_\beta = j, ..., Y_\psi = m, Z_\zeta = m - ... + (-1)^{N-1} j - (-1)^{N-1} i\big) \\
\leq & \sum_{i,j,...,m} \min\Big(P(A_\alpha = i, B_\beta = j, ..., Y_\psi = m), \\
& \qquad P\big(B_\beta = j, ..., Y_\psi = m, Z_\zeta = m - ... + (-1)^{N-1} j - (-1)^{N-1} i\big)\Big) \\
\leq & \min\Big(P(A_\alpha = a, B_\beta = b, ..., Y_\psi = y), \\
& \qquad P\big(B_\beta = b, ..., Y_\psi = y, Z_\zeta = y - ... + (-1)^{N-1} b - (-1)^{N-1} a\big)\Big) \\
& + \min\Big(\sum_{(i,j,...,m) \neq (a,b,...,y)} P(A_\alpha = a, B_\beta = b, ..., Y_\psi = y), \\
& \qquad \sum_{(i,j,...,m) \neq (a,b,...,y)} P\big(B_\beta = b, ..., Y_\psi = y, Z_\zeta = y - ... + (-1)^{N-1} b - (-1)^{N-1} a\big)\Big) \\
\equiv & \min\Big(P\big(A_\alpha = a, B_\beta = b, ..., Y_\psi = y\big), \\
& \qquad P\big(B_\beta = b, ..., Y_\psi = y, Z_\zeta = y - ... + (-1)^{N-1} q - (-1)^{N-1} a\big)\Big) \\
& + \min\Big(1 - P(A_\alpha = a, B_\beta = b, ..., Y_\psi = y), \\
& \qquad 1 - P\big(B_\beta = b, ..., Y_\psi = y, Z_\zeta = y - ... + (-1)^{N-1} b - (-1)^{N-1} a\big)\Big) \\
\equiv & 1 - \big|P(A_\alpha = a, B_\beta = b, ..., Y_\psi = y) \\
& \qquad - P\big(B_\beta = b, ..., Y_\psi = y, Z_\zeta = y - ... + (-1)^{N-1} b - (-1)^{N-1} a\big)\big|,
\end{aligned}
$$

for arbitrary a, b, ..., $y \in \{0, \ ... \ , d-1\}$. $\square$

A.5 Proof of condition (A.6)

Here we prove that, if for some setting $(\alpha', \alpha' + \beta' - 1, ..., \chi' + \psi' - 1)$ the highest probability $P(a^{max}_{(\alpha',...,\psi')}, ..., y^{max}_{(\alpha',...,\psi')} | \alpha', \alpha' + \beta' - 1, ..., \chi' + \psi' - 1)$ is bounded from below as in (A.1), then inequality (A.6) is true. Again, we proceed by *reductio ad absurdum*: Suppose (A.6) is false. Then,

$$
\begin{aligned}
& \big|P(A_{\alpha'} = a_{0(\alpha',...,\psi')}, ..., Y_{\chi'+\psi'-1} = y_{0(\alpha',...,\psi')}) \ - \\
& P(A_{\alpha'} = \dot{a}_{0(\alpha',...,\psi')}, ..., Y_{\chi'+\psi'-1} = \dot{y}_{0(\alpha',...,\psi')})\big| \ \leq \ \varepsilon,
\end{aligned} \tag{A.11}
$$

for all points $(a_{0(\alpha',...,\psi')}, ..., y_{0(\alpha',...,\psi')}) \in \mathcal{G}$, where $(\dot{a}_{0(\alpha',...,\psi')}, ..., \dot{y}_{0(\alpha',...,\psi')})$ is any other point of $\mathcal{G}$ whose distance

$\mathcal{D}(a_{0(\alpha',...,\psi')},...,y_{0(\alpha',...,\psi')};\dot{a}_{0(\alpha',...,\psi')},...,\dot{y}_{0(\alpha',...,\psi')})$ from $(a_{0(\alpha',...,\psi')},...,y_{0(\alpha',...,\psi')})$ is one:

$$\begin{aligned}
&\mathcal{D}(a_{0(\alpha',...,\psi')},...,y_{0(\alpha',...,\psi')};\dot{a}_{0(\alpha',...,\psi')},...,\dot{y}_{0(\alpha',...,\psi')})\\
&\equiv \mathcal{D}(a_{0(\alpha',...,\psi')};\dot{a}_{0(\alpha',...,\psi')})+...+\mathcal{D}(y_{0(\alpha',...,\psi')};\dot{y}_{0(\alpha',...,\psi')})\\
=\;& |a_{0(\alpha',...,\psi')}-\dot{a}_{0(\alpha',...,\psi')}|+...+|y_{0(\alpha',...,\psi')}-\dot{y}_{0(\alpha',...,\psi')}|=1 \qquad \text{(A.12)}
\end{aligned}$$

This, in turn, implies that

$$\begin{aligned}
&\sum_{a,...,y=0}^{d} P(A_{\alpha'}=a,...,Y_{\chi'+\psi'-1}=y)\\
\geq &\sum_{a,...,y=0}^{d} \big(P(A_{\alpha'}=a^{max}_{(\alpha',...,\psi')},...,Y_{\chi'+\psi'-1}=y^{max}_{(\alpha',...,\psi')})\\
&\qquad -\varepsilon\mathcal{D}(a,...,y;a^{max}_{(\alpha',...,\psi')},...,y^{max}_{(\alpha',...,\psi')})\big)\\
>&1+\frac{d^N(N-1)}{4}\varepsilon-\varepsilon(N-1)d^{N-2}\sum_{i=D^-_{(d)}}^{D^+_{(d)}}\mathcal{D}(i;0)\\
=&1+d^{N-2}(N-1)\varepsilon\left(\frac{d^2}{4}-\sum_{i=D^-_{(d)}}^{D^+_{(d)}}\mathcal{D}(i;0)\right),
\end{aligned} \tag{A.13}$$

where we have introduced $D^+_{(d)}=(d-1)/2=-D^-_{(d)}$, for d odd, and $D^+_{(d)}=d/2=-D^-_{(d)}+1$, for d even, and have used that $\mathcal{D}(a,...,y;a^{max}_{(\alpha',...,\psi')},...,y^{max}_{(\alpha',...,\psi')})\equiv\mathcal{D}(a;a^{max}_{(\alpha',...,\psi')})+...+\mathcal{D}(y;y^{max}_{(\alpha',...,\psi')})$. Notice that, for d odd, it is $\sum_{i=D^-_{(d)}}^{D^+_{(d)}}D(i;0)=(d^2-1)/4$, whereas for d even it is $\sum_{i=D^-_{(d)}}^{D^+_{(d)}}D(i;0)=d^2/4$. Thus, in both cases the last line of (A.13) is strictly greater than 1, which contradicts probability normalization. □

Appendix B

TOBL models and extensions

B.1 TOBL models for an arbitrary number of parties

Suppose that $M+N$ parties share a no-signaling set of correlations $P(A_1, ..., A_{M+N}|X_1, ..., X_{M+N})$. We are interested in which restrictions we should enforce over such a distribution in order to make sure that it cannot be used to violate a bipartite Bell inequality when parties $1, ..., M$ and $M+1, ..., M+N$ group together, even when several of such boxes are initially distributed.

One possibility is to demand the new bipartite object to behave as a generic classical bipartite device would. Viewed as bipartite, the distribution $P(A_1, ..., A_{M+N}|X_1, ..., X_{M+N})$ is such that it allows each of the two virtual parties (call them Alice and Bob) to perform sequential measurements on their subsystems. If we assume that the outcomes Alice and Bob observe are generated by a classical machine, it follows that $P(A_1, ..., A_{M+N}|X_1, ..., X_{M+N})$ can be written as:

$$P(a_1, ..., a_{M+N}|x_1, ..., x_{M+N}) = \sum_{\lambda} p_\lambda P_A^\lambda \cdot P_B^\lambda, \tag{B.1}$$

where we can regard each P_A^λ as a collection of probability distributions

$$P^\lambda_{\sigma(1)\to...\to\sigma(M)}(A_1, ..., A_M|X_1, ..., X_M), \tag{B.2}$$

one for each possible permutation σ of the M physical parties. Here $\sigma(1) \to ... \to \sigma(M)$ would indicate the process in which the first party to measure is $\sigma(1)$, followed by $\sigma(2)$, etc.

If, during a communication protocol, Alice must measure, say, x_3, she only has to choose an arbitrary permutation σ, with $\sigma(1) = 3$ and then generate a_3 according to the probability distribution $P^{\lambda}_{\sigma(1)\to...\to\sigma(M)}(A_1, ..., A_M|X_1, ..., X_M)$. If, at some time later, she needs to simulate the measurement of x_1 and $\sigma(2) \neq 1$, she would thus have to find a new permutation σ', with $\sigma'(1) = 3, \sigma'(2) = 1$, and generate a_1 from the conditional probability distribution $P^{\lambda}_{\sigma'(1)\to...\to\sigma'(M)}(a_1, a_2, a_4, ..., a_M|x_1, ..., x_M, a_3)$. By consistency, for any pair of permutations σ^1, σ^2 such that $\sigma^1(j) = \sigma^2(j)$, for all $j \in \{1, ..., m\}$, such distributions need to satisfy the condition:

$$\begin{aligned} &\textstyle\sum_{a>m} P^{\lambda}_{\sigma^1(1)\to...\to\sigma^1(M)}(a_1, ..., a_M|x_1...x_M) = \\ &= \textstyle\sum_{a>m} P^{\lambda}_{\sigma^2(1)\to...\to\sigma^2(M)}(a_1, ..., a_M|x_1...x_M), \end{aligned} \tag{B.3}$$

where $\sum_{a>m}$ denotes the sum over all variables $a_{\sigma(j)}$ with $j > m$. The same considerations apply for P^{λ}_B.

Local postselections on a prior sequence of Alice's and Bob's outcomes would imply changing the probabilities p_λ, but otherwise can be simulated in a similar fashion.

Putting everything together, we have that WCCPI operations over a set of (possibly different) boxes generate bipartite classical correlations if each box $P(A_1, ..., A_{N+M}|X_1, ..., X_{M+N})$, distributed along the partition $1...M|M+1...M+N$, admits a decomposition of the form

$$\begin{aligned} &P(a_1, ..., a_{M+N}|x_1, ..., x_{M+N}) \\ = &\sum_{\lambda} p_\lambda P^{\lambda}_{\sigma(1)\to...\to\sigma(M)}(a_{\sigma(1)}, ...,_{\sigma(M)}|x_{\sigma(1)}...x_{\sigma(M)}) \cdot \\ &\cdot \tilde{P}^{\lambda}_{\sigma'(M+1)\to...\to\sigma'(M+N)}(a_{\sigma'(M+1)}, ...,_{\sigma'(N+M)}|x_{\sigma'(N+1)}...x_{\sigma'(N+M)}). \end{aligned} \tag{B.4}$$

The reader can check that in the tripartite case the above description reduces to the TOBL definition given in the main text.

B.2 Probability distribution maximizing (5.14)

This appendix presents a tripartite non-signaling probability distribution that attains the maximum of 7/6 for the 'Guess Your Neighbor's Input' inequality, as well as its TOBL decomposition. To simplify notation, let us switch from $(a_1a_2a_3)$ to (abc); and from $(x_1x_2x_3)$, to (xyz). Now, consider the non-signaling tripartite probability distribution $P(A, B, C|X, Y, Z)$ given by the probabilities shown in Table B.1.

	000	001	010	011	100	101	110	111
000	$\frac{2}{3}$	0	0	0	0	0	0	$\frac{1}{3}$
001	$\frac{1}{3}$	$\frac{1}{3}$	0	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$
010	$\frac{1}{3}$	0	$\frac{1}{3}$	0	0	$\frac{1}{6}$	0	$\frac{1}{6}$
011	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$	0
100	$\frac{1}{3}$	0	0	$\frac{1}{6}$	$\frac{1}{3}$	0	0	$\frac{1}{6}$
101	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0
110	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0
111	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0

Table B.1: Tripartite probability distribution $P(A, B, C|X, Y, Z)$ attaining the maximum of 7/6 for the 'Guess Your Neighbor's Input' inequality, where the rows correspond to the inputs xyz and the columns to the outputs abc.

The value of the 'Guess Your Neighbor's Input' inequality for $P(A, B, C|X, Y, Z)$ equals

$$\mathbf{B}(P) = \frac{2}{3} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{7}{6} \neq 1, \tag{B.5}$$

and thus $P(A, B, C|X, Y, Z)$ cannot be approximated by any quantum system. Next we will prove that $P(A, B, C|X, Y, Z)$ belongs to the TOBL set of correlations, and so it is compatible with any bipartite information principle.

First, notice that $P(A, B, C|X, Y, Z)$ is invariant under permutations of the three parties. It is therefore enough to show that it admits a decomposition of the form (5.6) for the partition $A|BC$. Along this bipartition, probability distributions appearing in the decomposition (5.6) are such that the outcome a only depends on the measurement choice x for every given λ; let a_x denote this outcome for $x = 0, 1$. Conditions (5.7) and (5.8) tell us that for every λ the marginal $P_{B\to C}(B|Y, \lambda)$ is independent of z, and the marginal $P_{B\leftarrow C}(C|Z, \lambda)$ is independent of y. Thus, for $B \to C$ we have that b depends on y and c depends on both z and y. The possible outcomes will then be denoted b_y, c_{yz}. Similarly, for $B \leftarrow C$, the possible outcomes are b_{yz}, c_z. Tables B.2 and B.3 contain the output assignments corresponding to deterministic probability distributions together with the weights p_λ for $A|B \to C$ and $A|B \leftarrow C$, respectively. Note that, in agreement with (5.6), the outcome assignments for A and the weights p_λ are the same for both decompositions.

It is trivial to see that both tables indeed reproduce $P(A, B, C|X, Y, Z)$, and hence such a distribution belongs to the TOBL set.

λ	p_λ	a_0	a_1	b_0	b_1	c_{00}	c_{01}	c_{10}	c_{11}
1	1/12	0	0	0	1	0	1	0	1
2	1/12	0	0	0	0	0	1	0	1
3	1/12	0	0	0	0	0	0	0	1
4	1/12	0	0	0	1	0	0	0	1
5	1/12	0	1	0	1	0	0	0	0
6	1/12	0	1	0	0	0	1	0	0
7	1/12	0	1	0	0	0	0	0	0
8	1/12	0	1	0	1	0	1	0	0
9	1/6	1	0	1	1	1	1	1	0
10	1/6	1	1	1	0	1	0	1	1

Table B.2: TOBL decomposition into deterministic probabibility distributions characterized by outcome assignments for the bipartition $A|BC$ in the case $A|B \to C$. For every λ the outcome a only depends on x, and b only depends on y.

λ	p_λ	a_0	a_1	b_{00}	b_{01}	b_{10}	b_{11}	c_0	c_1
1	1/12	0	0	0	0	0	1	0	0
2	1/12	0	0	0	0	0	1	0	1
3	1/12	0	0	0	0	1	1	0	0
4	1/12	0	0	0	0	1	1	0	1
5	1/12	0	1	0	0	0	0	0	0
6	1/12	0	1	0	0	0	0	0	1
7	1/12	0	1	0	0	1	0	0	0
8	1/12	0	1	0	0	1	0	0	1
9	1/6	1	0	1	1	1	0	1	1
10	1/6	1	1	1	1	0	1	1	0

Table B.3: TOBL decomposition into deterministic probability distributions characterized by outcome assignments for the bipartition $A|BC$ in the case $A|B \leftarrow C$. For every λ the outcome a only depends on x, and c only depends on z.

Appendix C

Proof of full randomness amplification

Before entering the details of the proof of Theorem 6.2, let us introduce a convenient notation. In what follows, we sometimes treat conditional probability distributions as vectors. To avoid ambiguities, we explicitly label the vectors describing probability distributions with the arguments of the distributions in upper case. Thus, for example, we denote by $P(\mathbf{A}|\mathbf{X})$ the $(2^5 \times 2^5)$-dimensional vector with components $P(\mathbf{a}|\mathbf{x})$ for all $\mathbf{a}, \mathbf{x} \in \{0,1\}^5$. We also denote by I the vector with components $I(\mathbf{a}, \mathbf{x})$ given in (6.3). With this notation, inequality (6.2) can be written as the scalar product

$$I \cdot P(\mathbf{A}|\mathbf{X}) = \sum_{\mathbf{a},\mathbf{x}} I(\mathbf{a}, \mathbf{x}) P(\mathbf{a}|\mathbf{x}) \geq 6 \ .$$

Any probability distribution $P(\mathbf{a}|\mathbf{x})$ satisfies $C \cdot P(\mathbf{A}|\mathbf{X}) = 1$, where C is the vector with components $C(\mathbf{a}, \mathbf{x}) = 2^{-5}$. We also use this scalar-product notation for full blocks, as in

$$I^{\otimes N_d} \cdot P(B|Y) = \sum_{\mathbf{a}_1,\ldots\mathbf{a}_{N_d}} \sum_{\mathbf{x}_1,\ldots\mathbf{x}_{N_d}} \left[\prod_{i=1}^{N_d} I(\mathbf{a}_i, \mathbf{x}_i) \right] P(\mathbf{a}_1, \ldots \mathbf{a}_{N_d} | \mathbf{x}_1, \ldots \mathbf{x}_{N_d}) \ .$$

Following our upper/lower-case convention, the vector $P(B|Y, e, z)$ has components $P(b|y, e, z)$ for all b, y but fixed e, z.

The proof of Theorem 6.2 relies on two crucial lemmas, which are stated and proven in Sections C.2 and C.2.1, respectively. The first lemma bounds the distinguishability between the distribution distilled from a block of N_d

quintuplets and the ideal free random bit as function of the Bell violation (6.2) in each quintuplet. In particular, it guarantees that, if the correlations of all quintuplets in a given block violate inequality (6.2) sufficiently much, the bit distilled from the block will be indistinguishable from an ideal free random bit. The second lemma is required to guarantee that, if the statistics observed in all blocks but the distilling one are consistent with a maximal violation of inequality (6.2), the violation of the distilling block will be arbitrarily large.

C.1 Proof of Theorem 6.2

We begin with the identity

$$P(\text{guess}) = P(g=0)P(\text{guess}|g=0) + P(g=1)P(\text{guess}|g=1)\ . \tag{C.1}$$

As discussed, when the protocol is aborted ($g=0$) the distribution generated by the protocol and the ideal one are indistinguishable. In other words,

$$P(\text{guess}|g=0) = \frac{1}{2}\ . \tag{C.2}$$

If $P(g=0)=1$ then the protocol is secure, though in a trivial fashion. Next we address the non-trivial case where $P(g=1)>0$.

From formula (6.11), we have

$$\begin{aligned}
& P(\text{guess}|g=1) \\
= \ & \frac{1}{2} + \frac{1}{4}\sum_{k,\tilde{y},t} \max_z \sum_e \left| P(k,\tilde{y},t,e|z,g=1) - \frac{1}{2}P(\tilde{y},t,e|z,g=1) \right| \\
= \ & \frac{1}{2} + \frac{1}{4}\sum_{\tilde{y},t} P(\tilde{y},t|g=1) \sum_k \max_z \sum_e \left| P(k,e|z,\tilde{y},t,g=1) - \frac{1}{2}P(e|z,\tilde{y},t,g=1) \right| \\
\leq \ & \frac{1}{2} + \frac{1}{4}\sum_{\tilde{y},t} P(\tilde{y},t|g=1)\, 6\sqrt{N_d}\ (\alpha C + \beta I)^{\otimes N_d} \cdot P(\tilde{B}|\tilde{Y},t,g=1) \\
= \ & \frac{1}{2} + \frac{3\sqrt{N_d}}{2}\,(\alpha C + \beta I)^{\otimes N_d} \cdot \sum_{\tilde{y},t} P(\tilde{y},t|g=1) P(\tilde{B}|\tilde{Y},t,g=1) \\
= \ & \frac{1}{2} + \frac{3\sqrt{N_d}}{2}\,(\alpha C + \beta I)^{\otimes N_d} \cdot \sum_{t} P(t|g=1) P(\tilde{B}|\tilde{Y},t,g=1) \\
= \ & \frac{1}{2} + \frac{3\sqrt{N_d}}{2}\,(\alpha C + \beta I)^{\otimes N_d} \cdot \sum_{t} P(\tilde{B},t|\tilde{Y},g=1) \\
= \ & \frac{1}{2} + \frac{3\sqrt{N_d}}{2}\,(\alpha C + \beta I)^{\otimes N_d} \cdot P(\tilde{B}|\tilde{Y},g=1)
\end{aligned} \tag{C.3}$$

where the inequality is due to Lemma C.1 in Section C.2, we have used the no-signaling condition through $P(\tilde{y}, t|z, g = 1) = P(\tilde{y}, t|g = 1)$, in the second equality, and Bayes rule in the second and sixth equalities. From (C.3) and Lemma C.2 in Section C.2.1, we obtain

$$P(\text{guess}|g = 1) \;\leq\; \frac{1}{2} + \frac{3\sqrt{N_d}}{2}\left[\alpha^{N_d} + \frac{2\,N_b^{\log_2(1-\epsilon)}}{P(g=1)}\left(32\beta\epsilon^{-5}\right)^{N_d}\right] . \quad \text{(C.4)}$$

Finally, substituting bound (C.4) and equality (C.2) into (C.1), we obtain

$$P(\text{guess}) \;\leq\; \frac{1}{2} + \frac{3\sqrt{N_d}}{2}\left[P(g=1)\,\alpha^{N_d} + 2\,N_b^{\log_2(1-\epsilon)}\left(32\beta\epsilon^{-5}\right)^{N_d}\right] , \quad \text{(C.5)}$$

which, together with $P(g = 1) \leq 1$, implies (6.12).

C.2 Statement and proof of Lemma C.1

As mentioned, Lemma C.1 provides a bound on the distinguishability between the probability distribution obtained after distilling a block of N_d quintuplets and an ideal free random bit in terms of the Bell violation (6.2) in each quintuplet. The proof of Lemma C.1, in turn, requires two more lemmas, Lemma C.3 and Lemma C.4, stated and proven in Section C.2.2.

Lemma C.1. *For each integer $N_d \geq 130$ there exists a function $f : \{0,1\}^{N_d} \to \{0,1\}$ such that, for any given $(5N_d + 1)$-partite non-signaling distribution $P(\mathbf{a}_1, \ldots \mathbf{a}_{N_d}, e|\mathbf{x}_1, \ldots \mathbf{x}_{N_d}, z) = P(b, e|y, z)$, the random variable $k = f(\text{maj}(\mathbf{a}_1), \ldots \text{maj}(\mathbf{a}_{N_d}))$ satisfies*

$$\sum_k \max_z \sum_e \left|P(k, e|y, z) - \frac{1}{2}P(e|y, z)\right| \;\leq\; 6\sqrt{N_d}\,(\alpha C + \beta I)^{\otimes N_d} \cdot P(B|Y) \quad \text{(C.6)}$$

for all inputs $y = (\mathbf{x}_1, \ldots \mathbf{x}_{N_d}) \in \mathcal{X}^{N_d}$, and where α and β are real numbers such that $0 < \alpha < 1 < \beta$.

Proof of Lemma C.1. For any $\mathbf{x}_0 \in \mathcal{X}$ let $M_w^{\mathbf{x}_0}$ be the vector with components $M_w^{\mathbf{x}_0}(\mathbf{a}, \mathbf{x}) = \delta^w_{\text{maj}(\mathbf{a})}\delta^{\mathbf{x}_0}_{\mathbf{x}}$. The probability of getting $\text{maj}(\mathbf{a}) = w$ when using $\mathbf{x}_0$ as input can be written as $P(w|\mathbf{x}_0) = M_w^{\mathbf{x}_0} \cdot P(\mathbf{A}|\mathbf{X})$. Note that this probability can also be written as $P(w|\mathbf{x}_0) = \Gamma_w^{\mathbf{x}_0} \cdot P(\mathbf{A}|\mathbf{X})$, where $\Gamma_w^{\mathbf{x}_0} = M_w^{\mathbf{x}_0} + \Lambda_w^{\mathbf{x}_0}$ and $\Lambda_w^{\mathbf{x}_0}$ is any vector orthogonal to the no-signaling subspace, that is, such that $\Lambda_w^{\mathbf{x}_0} \cdot P(\mathbf{A}|\mathbf{X}) = 0$ for all non-signaling distribution

$P(\mathbf{A}|\mathbf{X})$. We can then write the left-hand side of (C.6) as

$$\begin{aligned}
&\sum_k \max_z \sum_e \left| P(k,e|y,z) - \frac{1}{2}P(e|y,z) \right| \\
= &\sum_k \max_z \sum_e P(e|y,z) \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) P(\mathbf{w}|y,e,z) \right| \\
= &\sum_k \max_z \sum_e P(e|z) \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \left(\bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right) \cdot P(B|Y,e,z) \right| \qquad \text{(C.7)}
\end{aligned}$$

where in the last equality we have used no-signaling through $P(e|y,z) = P(e|z)$ and the fact that the probability of obtaining the string of majorities $\mathbf{w}$ when inputting $y = (\mathbf{x}_1, \dots \mathbf{x}_{N_d}) \in \mathcal{X}^{N_d}$ can be written as

$$P(\mathbf{w}|y) = \left(\bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right) \cdot P(B|Y). \qquad \text{(C.8)}$$

In what follows, the absolute value of vectors is understood to be component-wise. Bound (C.7) can be rewritten as

$$\begin{aligned}
&\sum_k \max_z \sum_e \left| P(k,e|y,z) - \frac{1}{2}P(e|y,z) \right| \\
\leq &\sum_k \max_z \sum_e P(e|z) \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right| \cdot P(B|Y,e,z) \\
= &\sum_k \max_z \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right| \cdot \left(\sum_e P(e|z) P(B|Y,e,z) \right) \\
= &\sum_k \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right| \cdot P(B|Y), \qquad \text{(C.9)}
\end{aligned}$$

where the inequality follows from the fact that all the components of the vector $P(B|Y,e,z)$ are positive and no-signaling has been used again through $P(B|Y,z) = P(B|Y)$ in the last equality. The bound applies to any function f and holds for any choice of vectors $\Lambda^{\mathbf{x}_i}_w$ in $\Gamma^{\mathbf{x}_i}_w$. In what follows, we compute this bound for a specific choice of these vectors and function f.

Take $\Lambda^{\mathbf{x}_i}_w$ to be equal to the vectors $\Lambda^{\mathbf{x}_0}_w$ in Lemma C.3. These vectors then satisfy the bounds (C.21) and (C.30) in the same Lemma. Take f to be equal to the function whose existence is proven in Lemma C.4. Note

that the conditions needed for this Lemma to apply are satisfied because of bound (C.21) in Lemma C.3, and because the free parameter $N_d \geq 130$ satisfies $\left(3\sqrt{N_d}\right)^{-1/N_d} \geq \gamma = 0.9732$. With this choice of f and $\Lambda_w^{\mathbf{x}_i}$, bound (C.9) becomes

$$\begin{aligned} & \sum_k \max_z \sum_e \left| P(k,e|y,z) - \frac{1}{2}P(e|y,z) \right| \\ \leq \; & \sum_k 3\sqrt{N_d} \left(\bigotimes_{i=1}^{N_d} \Omega^{\mathbf{x}_i} \right) \cdot P(B|Y) \\ \leq \; & 6\sqrt{N_d}\,(\alpha C + \beta I)^{\otimes N_d} \cdot P(B|Y) \;, \end{aligned} \tag{C.10}$$

where we have used $\Omega^{\mathbf{x}_i} = \sqrt{(\Gamma_0^{\mathbf{x}_i})^2 + (\Gamma_1^{\mathbf{x}_i})^2}$, $\sum_k 3 = 6$, bound (C.21) in Lemma C.3 and bound (C.30) in Lemma C.4. □

C.2.1 Statement and proof of Lemma C.2

In this section we prove Lemma C.2. This Lemma bounds the Bell violation in the distillation block in terms of the probability of not aborting the protocol in step 4 and the number and size of the blocks, N_b and N_d.

Lemma C.2. *Let $P(b_1, \ldots b_{N_b}|y_1, \ldots y_{N_b})$ be a $(5N_dN_b)$-partite non-signaling distribution, $y_1, \ldots y_{N_b}$ and l the variables generated in steps 2 and 3 of the protocol, respectively, and α and β real numbers such that $0 < \alpha < 1 < \beta$; then*

$$(\alpha C + \beta I)^{\otimes N_d} \cdot P(\tilde{B}|\tilde{Y}, g=1) \;\leq\; \alpha^{N_d} + \frac{2\,N_b^{\log_2(1-\epsilon)}}{P(g=1)} \left(32\beta\epsilon^{-5}\right)^{N_d} \;. \tag{C.11}$$

Proof of Lemma C.2. According to definition (6.8) we have $I(\mathbf{a}_i, \mathbf{x}_i) \leq \delta^0_{r[b,y]}$ for all values of $b = (\mathbf{a}_1, \ldots \mathbf{a}_{N_d})$ and $y = (\mathbf{x}_1, \ldots \mathbf{x}_{N_d})$. This also implies $I(\mathbf{a}_i, \mathbf{x}_i) I(\mathbf{a}_j, \mathbf{x}_j) \leq \delta^0_{r[b,y]}$ and so on. Due to the property $0 < \alpha < 1 < \beta$, one has that $(\alpha\, 2^{-5})^{N_d - i}\beta^i \leq \beta^{N_d}$ for any $i = 1, \ldots N_d$. All this in

turn implies

$$
\begin{aligned}
& \prod_{i=1}^{N_d} \left[\alpha\, 2^{-5} + \beta I_i\right] \\
= \;& \left(\alpha\, 2^{-5}\right)^{N_d} + \left(\alpha\, 2^{-5}\right)^{N_d-1} \beta \sum_i I_i + \left(\alpha\, 2^{-5}\right)^{N_d-2} \beta^2 \sum_{i\neq j} I_i I_j + \cdots \\
\leq \;& \left(\alpha\, 2^{-5}\right)^{N_d} + \beta^{N_d} \left(\sum_i I_i + \sum_{i\neq j} I_i I_j + \cdots \right) \\
\leq \;& \left(\alpha\, 2^{-5}\right)^{N_d} + \beta^{N_d} \left(\sum_i \delta^0_{r[b,y]} + \sum_{i\neq j} \delta^0_{r[b,y]} + \cdots \right) \\
\leq \;& \left(\alpha\, 2^{-5}\right)^{N_d} + \beta^{N_d} \left(2^{N_d} - 1\right) \delta^0_{r[b,y]} \;\leq\; \left(\alpha\, 2^{-5}\right)^{N_d} + (\beta\, 2)^{N_d}\, \delta^0_{r[b,y]} \quad \text{(C.12)}
\end{aligned}
$$

where $I_i = I(\mathbf{a}_i, \mathbf{x}_i)$. This implies that

$$
\begin{aligned}
& (\alpha C + \beta I)^{\otimes N_d} \cdot P(B|Y, g=1) \\
= \;& \sum_{\mathbf{a}_1,\ldots\mathbf{a}_{N_d}} \sum_{\mathbf{x}_1,\ldots\mathbf{x}_{N_d}} \prod_{i=1}^{N_d} \left[\alpha\, 2^{-5} + \beta I(\mathbf{a}_i, \mathbf{x}_i)\right] P(\mathbf{a}_1, \ldots \mathbf{a}_{N_d} | \mathbf{x}_1, \ldots \mathbf{x}_{N_d}, g=1) \\
\leq \;& \sum_{b,y} \left[\left(\alpha\, 2^{-5}\right)^{N_d} + (2\beta)^{N_d} \delta^0_{r[b,y]} \right] P(b|y, g=1) \\
= \;& \alpha^{N_d} \sum_y 2^{-5N_d} + (2\beta)^{N_d} \sum_y P(r=0|y, g=1) \\
= \;& \alpha^{N_d} + (2\beta)^{N_d} \sum_y P(r=0|y, g=1) \\
= \;& \alpha^{N_d} + (2\beta)^{N_d} \sum_y \frac{P(r=0, y|g=1)}{P(y|g=1)} \;. \quad \text{(C.13)}
\end{aligned}
$$

We can now bound $P(y|g=1)$ taking into account that y denotes a $5N_d$-bit string generated by the ϵ-source $\mathcal{S}$ that remains after step 2 in the protocol. Note that only half of the 32 possible 5-bit inputs $\mathbf{x}$ generated by the source belong to $\mathcal{X}$ and remain after step 2. Thus, $P((\mathbf{x}_1, \ldots, \mathbf{x}_{N_d}) \in \mathcal{X}^{N_d} | g=1) \leq 16^{N_d}(1-\epsilon)^{5N_d}$, where we used (6.7). This, together with $P((\mathbf{x}_1, \ldots, \mathbf{x}_{N_d})|g=1) \geq \epsilon^{5N_d}$ implies that

$$
P(y|g=1) \geq \left(\frac{\epsilon^5}{16(1-\epsilon)^5} \right)^{N_d} . \quad \text{(C.14)}
$$

Substituting this bound in (C.13), and summing over y, gives

$$(\alpha C + \beta I)^{\otimes N_d} \cdot P(B|Y, g=1) \;\leq\; \alpha^{N_d} + (2\beta)^{N_d} \left(\frac{16(1-\epsilon)^5}{\epsilon^5} \right)^{N_d} P(r=0|g=1)\,. \tag{C.15}$$

In what follows we use the notation

$$P(1_1, 0_2, 1_3, 1_4, \ldots) = P(r[b_1, y_1] = 1, r[b_2, y_2] = 0, r[b_3, y_3] = 1, r[b_4, y_4] = 1, \ldots)\,.$$

According to (6.9), the protocol aborts ($g = 0$) if there is at least a "not right" block ($r[b_j, y_j] = 0$ for some $j \neq l$). While abortion also happens if there are more than one "not right" block, in what follows we lower-bound $P(g = 0)$ by the probability that there is only one "not right" block:

$$\begin{aligned}
1 \;&\geq\; P(g=0) \\
&\geq\; \sum_{l=1}^{N_b} P(l) \sum_{l'=1, l' \neq l}^{N_b} P(1_1, \ldots 1_{l-1}, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&\geq\; \sum_{l} P(l) \sum_{l' \neq l} P(1_1, \ldots 1_{l-1}, 1_l, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&=\; \sum_{l'} \left[\textstyle\sum_{l \neq l'} P(l) \right] P(1_1, \ldots 1_{l-1}, 1_l, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&=\; \sum_{l'} [1 - P(l')]\, P(1_1, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}),
\end{aligned} \tag{C.16}$$

where, when performing the sum over l, we have used that $P(1_1, \ldots 1_{l-1}, 1_l, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \equiv P(1_1, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b})$ does not depend on l. Bound (6.7) implies

$$\frac{1 - P(l)}{P(l)} \;\geq\; \frac{1 - (1-\epsilon)^{\log_2 N_b}}{(1-\epsilon)^{\log_2 N_b}} \;=\; N_b^{\log_2 \frac{1}{1-\epsilon}} - 1 \;\geq\; \frac{N_b^{\log_2 \frac{1}{1-\epsilon}}}{2}\,, \tag{C.17}$$

where the last inequality holds for sufficiently large N_b. Using this and (C.16), we obtain

$$\begin{aligned}
1 \;&\geq\; \frac{1}{2} \sum_{l'} N_b^{\log_2 \frac{1}{1-\epsilon}}\, P(l')\, P(1_1, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&\geq\; \frac{1}{2}\, N_b^{\log_2 \frac{1}{1-\epsilon}}\, P(\tilde{r} = 0, g = 1)\,,
\end{aligned} \tag{C.18}$$

where $\tilde{r} = r[b_l, y_l]$. This together with (C.15) implies

$$\begin{aligned} & (\alpha C + \beta I)^{\otimes N_d} \cdot P(\tilde{B}|\tilde{Y}, g = 1) \\ \leq \; & \alpha^{N_d} + (2\beta)^{N_d} \left(\frac{16(1-\epsilon)^5}{\epsilon^5} \right)^{N_d} P(\tilde{r} = 0|g = 1) \qquad \text{(C.19)} \\ \leq \; & \alpha^{N_d} + \frac{2}{P(g=1)} \left(\frac{32\beta(1-\epsilon)^5}{\epsilon^5} \right)^{N_d} N_b^{\log_2(1-\epsilon)} , \qquad \text{(C.20)} \end{aligned}$$

where, in the second inequality, Bayes rule was again invoked. Inequality (C.20), in turn, implies (C.11). □

C.2.2 Statement and proof of the additional Lemmas

Lemma C.3. *For each* $\mathbf{x}_0 \in \mathcal{X}$ *there are three vectors* $\Lambda_0^{\mathbf{x}_0}, \Lambda_1^{\mathbf{x}_0}, \Lambda_2^{\mathbf{x}_0}$ *orthogonal to the non-signaling subspace such that for all* $w \in \{0,1\}$ *and* $\mathbf{a}, \mathbf{x} \in \{0,1\}^5$ *they satisfy*

$$\begin{aligned} & \sqrt{[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]^2 + [M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]^2} \\ \leq \; & \alpha C(\mathbf{a},\mathbf{x}) + \beta I(\mathbf{a},\mathbf{x}) + \Lambda_2^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) \qquad \text{(C.21)} \end{aligned}$$

and

$$\begin{aligned} & |M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})| \\ \leq \; & \gamma \sqrt{[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]^2 + [M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]^2} \qquad \text{(C.22)} \end{aligned}$$

where $\alpha = 0.8842$, $\beta = 1.260$ *and* $\gamma = 0.9732$.

Proof of Lemma C.3. The proof of this lemma is numeric but rigorous. It is based on two linear-programming minimization problems, which are carried for each value of $\mathbf{x}_0 \in \mathcal{X}$. We have repeated this process for different values of γ, finding that $\gamma = 0.9732$ is roughly the smallest value for which the linear-programs described below are feasible.

The fact that the vectors $\Lambda_0^{\mathbf{x}_0}, \Lambda_1^{\mathbf{x}_0}, \Lambda_2^{\mathbf{x}_0}$ are orthogonal to the non-signaling subspace can be written as linear equalities

$$D \cdot \Lambda_w^{\mathbf{x}_0} = \mathbf{0} \qquad \text{(C.23)}$$

for $w \in \{0,1,2\}$, where $\mathbf{0}$ is the zero vector and D is a matrix whose rows constitute a basis of non-signaling probability distributions. A geometrical interpretation of constraint (C.21) is that the point in the plane with coordinates $[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}), M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})] \in \mathbb{R}^2$ is inside a circle

of radius $\alpha C(\mathbf{a},\mathbf{x}) + \beta I(\mathbf{a},\mathbf{x}) + \Lambda_2^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})$ centered at the origin. All points inside an octagon inscribed in this circle also satisfy constraint (C.21). The points of such an inscribed octagon are the ones satisfying the following set of linear constraints:

$$\begin{aligned} & [M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]\,\eta\cos\theta + [M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]\,\eta\sin\theta \\ \leq\ & \alpha C(\mathbf{a},\mathbf{x}) + \beta I(\mathbf{a},\mathbf{x}) + \Lambda_2^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\ , \end{aligned} \tag{C.24}$$

for all $\theta \in \{\frac{\pi}{8}, \frac{3\pi}{8}, \frac{5\pi}{8}, \frac{7\pi}{8}, \frac{9\pi}{8}, \frac{11\pi}{8}, \frac{13\pi}{8}, \frac{15\pi}{8}\}$, where $\eta = (\cos\frac{\pi}{8})^{-1} \approx 1.082$. In other words, the eight conditions (C.24) imply constraint (C.21). From now on, we only consider these eight linear constraints (C.24). With a bit of algebra, one can see that inequality (C.22) is equivalent to the two almost linear inequalities there was an error in the following equation, as the prefactor in terms of γ was wrong. Please check what was computed and how it affects to γ and, then, to the value of N_d

$$\pm\,[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]\ \leq\ \sqrt{\frac{\gamma^2}{1-\gamma^2}}\,|M_{\bar{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_{\bar{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})|\ , \tag{C.25}$$

for all $w \in \{0,1\}$, where $\bar{w} = 1 - w$. Clearly, the problem is not linear because of the absolute values. The computation described in what follows constitutes a trick to make a good guess for the signs of the terms in the absolute value of (C.25), so that the problem can be made linear by adding extra constraints.

The first computational step consists of a linear-programming minimization of α subject to the constraints (C.23), (C.24), where the minimization is performed over the variables $\alpha, \beta, \Lambda_0^{\mathbf{x}_0}, \Lambda_1^{\mathbf{x}_0}, \Lambda_2^{\mathbf{x}_0}$. This step serves to guess the signs

$$\sigma_w(\mathbf{a},\mathbf{x})\ =\ \mathrm{sign}[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]\ , \tag{C.26}$$

for all $w, \mathbf{a}, \mathbf{x}$, where the value of $\Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})$ corresponds to the solution of the above minimization. Once we have identified all these signs, we can write the inequalities (C.25) in a linear fashion:

$$\sigma_w(\mathbf{a},\mathbf{x})\,[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})] \geq 0\ , \tag{C.27}$$

$$\sigma_w(\mathbf{a},\mathbf{x})\,[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})] \leq \sqrt{\frac{\gamma^2}{1-\gamma^2}}\,\sigma_{\bar{w}}(\mathbf{a},\mathbf{x})\,[M_{\bar{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_{\bar{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})]\ , \tag{C.28}$$

for all $w \in \{0,1\}$.

The second computational step consists of a linear-programming minimization of α subjected to the constraints (C.23), (C.24), (C.27), (C.28), over the variables $\alpha, \beta, \Lambda_0^{\mathbf{x}_0}, \Lambda_1^{\mathbf{x}_0}, \Lambda_2^{\mathbf{x}_0}$. Clearly, any solution to this problem is also a solution to the original formulation of the Lemma. The minimization was performed for any $\mathbf{x}_0 \in \mathcal{X}$ and the values of α, β turned out to be independent of $\mathbf{x}_0 \in \mathcal{X}$. These obtained numerical values are the ones appearing in the formulation of the Lemma. □

Note that Lemma C.3 allows one to bound the predictability of maj($\mathbf{a}$) by a linear function of the 5-party Mermin violation. This can be seen by computing $\Gamma_w^{\mathbf{x}_0} \cdot P(\mathbf{A}|\mathbf{X})$ and applying the bounds in the Lemma. In principle, one expects this bound to exist, as the predictability is smaller than one at the point of maximal violation, as proven in Theorem 6.1, and equal to one at the point of no violation. However, we were unable to find it. This is why we had to resort to the linear optimization technique given above, which moreover provides the bounds (C.21) and (C.22) necessary for the security proof.

Lemma C.4. *Let N_d be a positive integer and let $\Gamma_w^i(\mathbf{a}, \mathbf{x})$ be a given set of real coefficients such that for all $i \in \{1, \dots N_d\}$, $w \in \{0, 1\}$ and $\mathbf{a}, \mathbf{x} \in \{0, 1\}^5$ they satisfy*

$$\left|\Gamma_w^i(\mathbf{a}, \mathbf{x})\right| \leq \left(3\sqrt{N_d}\right)^{-1/N_d} \Omega_i(\mathbf{a}, \mathbf{x}) \ , \tag{C.29}$$

where $\Omega_i(\mathbf{a}, \mathbf{x}) = \sqrt{\Gamma_0^i(\mathbf{a}, \mathbf{x})^2 + \Gamma_1^i(\mathbf{a}, \mathbf{x})^2}$. There exists a function $f : \{0, 1\}^{N_d} \to \{0, 1\}$ such that for each sequence $(\mathbf{a}_1, \mathbf{x}_1), \dots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ we have

$$\left|\sum_{\mathbf{w}} \left(\delta_{f(\mathbf{w})}^k - \frac{1}{2}\right) \prod_{i=1}^{N_d} \Gamma_{w_i}^i(\mathbf{a}_i, \mathbf{x}_i)\right| \ \leq \ 3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i(\mathbf{a}_i, \mathbf{x}_i) \ , \tag{C.30}$$

where the sum runs over all $\mathbf{w} = (w_1, \dots w_{N_d}) \in \{0, 1\}^{N_d}$.

Proof of Lemma (C.4)**.** First, note that for a sequence $(\mathbf{a}_1, \mathbf{x}_1), \dots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ for which there is at least one value of $i \in \{1, \dots N_d\}$ satisfying $\Gamma_0^i(\mathbf{a}_i, \mathbf{x}_i) = \Gamma_1^i(\mathbf{a}_i, \mathbf{x}_i) = 0$, both the left-hand side and the right-hand side of (C.30) are equal to zero, hence, inequality (C.30) is satisfied independently of the function f. Therefore, in what follows, we only consider sequences $(\mathbf{a}_1, \mathbf{x}_1), \dots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ for which either $\Gamma_0^i(\mathbf{a}_i, \mathbf{x}_i) \neq 0$ or $\Gamma_1^i(\mathbf{a}_i, \mathbf{x}_i) \neq 0$, for all $i = 1, \dots N_d$. Or, equivalently, we consider sequences such that

$$\prod_{i=1}^{N_d} \Omega_i(\mathbf{a}_i, \mathbf{x}_i) > 0 \ . \tag{C.31}$$

The existence of the function f satisfying (C.30) for all such sequences is shown with a probabilistic argument. We consider the situation where f is picked from the set of all functions mapping $\{0,1\}^{N_d}$ to $\{0,1\}$ with uniform probability, and upper-bound the probability that the chosen function does not satisfy the constraint (C.30) for all k and all sequences $(\mathbf{a}_1, \mathbf{x}_1), \ldots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ satisfying (C.31). This upper bound is shown to be smaller than one. Therefore there must exist at least one function satisfying (C.30).

For each $\mathbf{w} \in \{0,1\}^{N_d}$ consider the random variable $F_{\mathbf{w}} = (\delta^0_{f(\mathbf{w})} - \frac{1}{2}) \in \{\frac{1}{2}, -\frac{1}{2}\}$, where f is picked from the set of all functions mapping $\{0,1\}^{N_d} \to \{0,1\}$ with uniform distribution. This is equivalent to saying that the 2^{N_d} random variables $\{F_{\mathbf{w}}\}_{\mathbf{w}}$ are independent and identically distributed according to $\Pr\{F_{\mathbf{w}} = \pm\frac{1}{2}\} = \frac{1}{2}$. For ease of notation, let us fix a sequence $(\mathbf{a}_1, \mathbf{x}_1), \ldots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ satisfying (C.31) and use the short-hand notation $\Gamma^i_{w_i} = \Gamma^i_{w_i}(\mathbf{a}_i, \mathbf{x}_i)$.

We proceed using the same ideas as in the derivation of the exponential Chebyshev's Inequality. For any $\mu, \nu \geq 0$, we have

$$
\begin{aligned}
& \Pr\left\{\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq \mu\right\} \\
= \; & \Pr\left\{\nu\left(-\mu + \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i}\right) \geq 0\right\} \\
= \; & \Pr\left\{\exp\left(-\nu\mu + \nu \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i}\right) \geq 1\right\} \\
\leq \; & \mathbb{E}\left[\exp\left(-\nu\mu + \nu \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i}\right)\right] && \text{(C.32)} \\
= \; & \mathbb{E}\left[\mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \exp\left(\nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i}\right)\right] \\
= \; & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \mathbb{E}\left[\exp\left(\nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i}\right)\right] && \text{(C.33)} \\
\leq \; & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \mathbb{E}\left[1 + \nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} + \left(\nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i}\right)^2\right] . && \text{(C.34)}
\end{aligned}
$$

Here $\mathbb{E}$ stands for the average over all $F_{\mathbf{w}}$. In (C.32) we have used that any positive random variable X satisfies $\Pr\{X \geq 1\} \leq \mathbb{E}[X]$. In (C.33) we have

used that the $\{F_{\mathbf{w}}\}_{\mathbf{w}}$ are independent. Finally, in (C.34) we have used that $\mathrm{e}^{\eta} \leq 1 + \eta + \eta^2$, which is only valid if $\eta \leq 1$. Therefore, we must show that

$$\left| \frac{\nu}{2} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right| \leq 1, \tag{C.35}$$

which is done below, when setting the value of ν. In what follows we use the chain of inequalities (C.34), the fact that $\mathbb{E}[F_{\mathbf{w}}] = 0$ and $\mathbb{E}[F_{\mathbf{w}}^2] = 1/4$, bound $1 + \eta \leq \mathrm{e}^{\eta}$ for $\eta \geq 0$, and the definition $\Omega_i^2 = (\Gamma_0^i)^2 + (\Gamma_1^i)^2$:

$$\begin{aligned}
& \Pr\left\{ \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq \mu \right\} \\
\leq\ & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \left(1 + \mathbb{E}[F_{\mathbf{w}}]\, \nu \prod_{i=1}^{N_d} \Gamma^i_{w_i} + \mathbb{E}[F_{\mathbf{w}}^2]\, \nu^2 \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
=\ & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \left(1 + \frac{\nu^2}{4} \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
\leq\ & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \exp\left(\frac{\nu^2}{4} \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
=\ & \exp\left(-\nu\mu + \sum_{\mathbf{w}} \frac{\nu^2}{4} \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
=\ & \exp\left(-\nu\mu + \frac{\nu^2}{4} \prod_{i=1}^{N_d} \Omega_i^2 \right)
\end{aligned} \tag{C.36}$$

In order to optimize this upper bound, we minimize the exponent over ν. This is done by differentiating with respect to ν and equating to zero, which gives

$$\nu = 2\,\mu \prod_{i=1}^{N_d} \Omega_i^{-2} \ . \tag{C.37}$$

Note that constraint (C.31) implies that the inverse of Ω_i exists. Since we assume $\mu \geq 0$, the initial assumption $\nu \geq 0$ is satisfied by the solution (C.37). By substituting (C.37) in (C.36) and rescaling the free parameter μ as

$$\tilde{\mu} = \frac{\mu}{\prod_{i=1}^{N_d} \Omega_i} \ , \tag{C.38}$$

we obtain

$$\Pr\left\{\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq \tilde{\mu} \prod_{i=1}^{N_d} \Omega_i\right\} \leq \mathrm{e}^{-\tilde{\mu}^2} \ , \tag{C.39}$$

for any $\tilde{\mu} \geq 0$ consistent with condition (C.35). We now choose $\tilde{\mu} = 3\sqrt{N_d}$, see Eq. (C.30), getting

$$\Pr\left\{\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq 3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i\right\} \leq \mathrm{e}^{-9N_d} \ . \tag{C.40}$$

With this assignment, and using (C.37) and (C.38), condition (C.35), yet to be fulfilled, becomes

$$3\sqrt{N_d} \prod_{i=1}^{N_d} \frac{|\Gamma^i_{w_i}|}{\Omega_i} \leq 1 \ , \tag{C.41}$$

which now holds because of the initial premise (C.29).

Bound (C.40) applies to each of the sequences $(\mathbf{a}_1, \mathbf{x}_1), \ldots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ satisfying (C.31), and there are at most 4^{5N_d} of them. Hence, the probability that the random function f does not satisfy the bound

$$\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq 3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i, \tag{C.42}$$

for at least one of such sequences, is at most $4^{5N_d}\mathrm{e}^{-9N_d}$, which is smaller than 1/2 for any value of N_d. A similar argument proves that the probability that the random function f does not satisfy the bound

$$\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \leq -3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i, \tag{C.43}$$

for at least one sequence satisfying (C.31) is also smaller than 1/2. The lemma now easily follows from these two results. □

C.3 Final remarks

The main goal was to prove full randomness amplification. In this Appendix, we have shown how our protocol, based on quantum non-local correlations, achieves this task. Unfortunately, we are not able to provide an explicit

description of the function $f : \{0,1\}^{N_d} \to \{0,1\}$ which maps the outcomes of the black boxes to the final random bit k; we merely show its existence. Such function may be obtained through an algorithm that searches over the set of all functions until it finds one satisfying (C.30). The problem with this method is that the set of all functions has size 2^{N_d}, which makes the search computationally costly. However, this problem can be fixed by noticing that the random choice of f in the proof of Lemma C.4 can be restricted to a four-universal family of functions, with size polynomial in N_d. This observation will be developed in future work.

A more direct approach could consist of studying how the randomness in the measurement outcomes for correlations maximally violating the Mermin inequality increases with the number of parties. We solved linear optimization problems similar to those used in Theorem 6.1 which showed that for 7 parties Eve's predictability is $2/3$ for a function of 5 bits defined by $f(00000) = 0$, $f(01111) = 0$, $f(00111) = 0$ and $f(\mathbf{x}) = 1$ otherwise. Note that this value is lower than the earlier $3/4$ and also that the function is different from the majority-vote. We were however unable to generalize these results for an arbitrary number of parties, which forced us to adopt a less direct approach. Note in fact that our protocol can be interpreted as a huge multipartite Bell test from which a random bit is extracted by classical processing of some of the measurement outcomes.

We conclude by stressing again that the reason why randomness amplification becomes possible using non-locality is because the randomness certification is achieved by a Bell inequality violation. There already exist several protocols, both in classical and quantum information theory, in which imperfect randomness is processed to generate perfect (or arbitrarily close to perfect) randomness. However, all these protocols, e.g. two-universal hashing or randomness extractors, always require additional good-quality randomness to perform such distillation. On the contrary, if the initial imperfect randomness has been certified by a Bell inequality violation, the distillation procedure can be done with a deterministic hash function (see [Mas09] or Lemma C.1 above). This property makes Bell-certified randomness fundamentally different from any other form of randomness, and is the key for the success of our protocol.

Bibliography

[ABB+10] Mafalda L. Almeida, Jean-Daniel Bancal, Nicolas Brunner, Antonio Acín, Nicolas Gisin, and Stefano Pironio. Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage. *Phys. Rev. Lett.*, 104(23):230404–, June 2010.

[ABCB12] Johan Ahrens, Piotr Badziag, Adan Cabello, and Mohamed Bourennane. Experimental device-independent tests of classical and quantum dimensions. *Nat Phys*, 8(8):592–595, August 2012.

[ABPS09] Jonathan Allcock, Nicolas Brunner, Marcin Pawlowski, and Valerio Scarani. Recovering part of the boundary between quantum and nonquantum correlations from information causality. *Phys. Rev. A*, 80(4):040103–, October 2009.

[ABSW92] L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman. Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes. *Phys. Rev. A*, 45(11):8185–8189, June 1992.

[ACSA10] Mafalda L. Almeida, Daniel Cavalcanti, Valerio Scarani, and Antonio Acín. Multipartite fully nonlocal quantum states. *Phys. Rev. A*, 81(5):052111–, May 2010.

[ADGL02] A. Acín, T. Durt, N. Gisin, and J. I. Latorre. Quantum nonlocality in two three-level systems. *Phys. Rev. A*, 65(5):052325–, May 2002.

[ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of bell's inequalities using time- varying analyzers. *Phys. Rev. Lett.*, 49(25):1804–1807, December 1982.

[AKR+10] Ali Ahanj, Samir Kunkri, Ashutosh Rai, Ramij Rahaman, and Pramod S. Joag. Bound on hardy's nonlocality from the principle of information causality. *Phys. Rev. A*, 81(3):032103–, March 2010.

[AMP12] A. Acín, Serge Massar, and Stefano Pironio. Randomness versus nonlocality and entanglement. *Phys. Rev. Lett.*, 108(10):100402–, March 2012.

[Ard92] M. Ardehali. Bell inequalities with a magnitude of violation that grows exponentially with the number of particles. *Phys. Rev. A*, 46(9):5375–5378, November 1992.

[BBB+12] Jean-Daniel Bancal, Cyril Branciard, Nicolas Brunner, Nicolas Gisin, and Yeong-Cherng Liang. A framework for the study of symmetric full-correlation bell-like inequalities. *Journal of Physics A: Mathematical and Theoretical*, 45(12):125301–, 2012.

[BBGL11] Jean-Daniel Bancal, Nicolas Brunner, Nicolas Gisin, and Yeong-Cherng Liang. Detecting genuine multipartite quantum nonlocality: A simple approach and generalization to arbitrary dimensions. *Phys. Rev. Lett.*, 106(2):020405–, January 2011.

[BBL+06] Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Phys. Rev. Lett.*, 96(25):250401–, June 2006.

[BBT05] Gilles Brassard, Anne Broadbent, and Alain Tapp. Quantum pseudo-telepathy, 2005.

[BC90] Samuel L Braunstein and Carlton M Caves. Wringing out better bell inequalities. *Annals of Physics*, 202(1):22–56, August 1990.

[BCH+02] Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu. Quantum nonlocality, bell inequalities, and the memory loophole. *Phys. Rev. A*, 66(4):042111–, October 2002.

[BCT99] Gilles Brassard, Richard Cleve, and Alain Tapp. Cost of exactly simulating quantum entanglement with classical communication. *Phys. Rev. Lett.*, 83(9):1874–1877, August 1999.

[Bel64] J. S. Bell. *Physics*, 1:195, 1964.

[BG10] J. Barrett and N. Gisin. How much measurement independence is needed in order to demonstrate nonlocality?, 2010.

[BHK05] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95(1):010503–, June 2005.

[BKP06] Jonathan Barrett, Adrian Kent, and Stefano Pironio. Maximally nonlocal and monogamous quantum correlations. *Phys. Rev. Lett.*, 97(17):170409–4, October 2006.

[Boh52] David Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. i. *Phys. Rev.*, 85(2):166–179, January 1952.

[BP05] Jonathan Barrett and Stefano Pironio. Popescu-rohrlich correlations as a unit of nonlocality. *Phys. Rev. Lett.*, 95(14):140401–, September 2005.

[BPA+08] Nicolas Brunner, Stefano Pironio, Antonio Acin, Nicolas Gisin, André Allan Méthot, and Valerio Scarani. Testing the dimension of hilbert spaces. *Phys. Rev. Lett.*, 100(21):210503–, May 2008.

[BPBG11] J. Barrett, S. Pironio, J-D. Bancal, and N. Gisin. The definition of multipartite nonlocality, 2011.

[BS09] Nicolas Brunner and Paul Skrzypczyk. Nonlocality distillation and postquantum theories with trivial communication complexity. *Phys. Rev. Lett.*, 102(16):160403–, April 2009.

[BT09] H. Briët, J. Buhrman and B. Toner. A generalized grothendieck inequality and entanglement in xor games, 2009.

[BV] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.*

[BvDHT99] Harry Buhrman, Wim van Dam, Peter Høyer, and Alain Tapp. Multiparty quantum communication complexity. *Phys. Rev. A*, 60(4):2737–2741, October 1999.

[BW92] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Phys. Rev. Lett.*, 69(20):2881–2884, November 1992.

[Cab01] Adán Cabello. "all versus nothing" inseparability for two observers. *Phys. Rev. Lett.*, 87(1):010403–, June 2001.

[Can01] R. Canetti. Universally composable security: a new paradigm for cryptographic protocols. *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 136–145, 2001.

[CBH03] Rob Clifton, Jeffrey Bub, and Hans Halvorson. Characterizing quantum theory in terms of information-theoretic constraints. *Foundations of Physics*, 33(11):1561–1591, November 2003.

[CBP+05] C. Cinelli, M. Barbieri, R. Perris, P. Mataloni, and F. De Martini. All-versus-nothing nonlocality test of quantum mechanics by two-photon hyperentanglement. *Phys. Rev. Lett.*, 95(24):240405–, December 2005.

[CEGA96] Adán Cabello, JoséM. Estebaranz, and Guillermo García-Alcaine. Bell-kochen-specker theorem: A proof with 18 vectors. *Physics Letters A*, 212(4):183–187, March 1996.

[CGL+02] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. Bell inequalities for arbitrarily high-dimensional systems. *Phys. Rev. Lett.*, 88(4):040404–, January 2002.

[CH74] John F. Clauser and Michael A. Horne. Experimental consequences of objective local theories. *Phys. Rev. D*, 10(2):526–535, July 1974.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(15):880–884, October 1969.

[CK06] John Conway and Simon Kochen. The free will theorem, 2006.

[Col07] R. Colbeck. *Quantum and Relativistic Protocols for Secure Multi-Party Computation*. PhD thesis, Univ. of Cambridge, 2007.

[CR08] Roger Colbeck and Renato Renner. Hidden variable models for quantum theory cannot have any local part. *Phys. Rev. Lett.*, 101(5):050403–4, August 2008.

[CR12] Roger Colbeck and Renato Renner. Free randomness can be amplified. *Nat Phys*, 8(6):450–454, June 2012.

[CSS10] Daniel Cavalcanti, Alejo Salles, and Valerio Scarani. Macroscopically local correlations can violate information causality. *Nat Commun*, 1:136–, December 2010.

[CVB+10] A. Chiuri, G. Vallone, N. Bruno, C. Macchiavello, D. Bruß, and P. Mataloni. Hyperentangled mixed phased dicke states: Optical design and detection. *Phys. Rev. Lett.*, 105(25):250501–, December 2010.

[Dam00] W. V. Dam. *Nonlocality and communication complexity*. PhD thesis, University of Oxford, 2000.

[Dam05] W. V. Dam. Implausible consequences of superstrong nonlocality, 2005.

[DAPG+12] M. Dall' Arno, E. Passaro, R. Gallego, M. Pawlowski, and A. Acín. Attacks on semi-device independent quantum protocols, 2012.

[DAPGA12] Michele Dall' Arno, Elsa Passaro, Rodrigo Gallego, and Antonio Acín. Robustness of device-independent dimension witnesses. *Phys. Rev. A*, 86(4):042312–, October 2012.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[EPR92] Avshalom C. Elitzur, Sandu Popescu, and Daniel Rohrlich. Quantum nonlocality for each pair in an ensemble. *Physics Letters A*, 162(1):25–28, January 1992.

[FGS11] S. Fehr, R. Gelles, and C. Schaffner. Security and composability of randomness expansion from bell inequalities, 2011.

[Fin82] Arthur Fine. Hidden variables, joint probability, and the bell inequalities. *Phys. Rev. Lett.*, 48(5):291–, February 1982.

[Gal09] Ernesto F. Galvão. Economical ontological models for discrete quantum systems. *Phys. Rev. A*, 80(2):022106–, August 2009.

[GHZ89] D. M. Greenberger, M. A. Horne, and A. Zeilinger. *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*. Kluwer, 1989.

[HA07] T. Harrigan, N. Rudolph and S. Aaronson. Representing probabilistic data via ontological models, 2007.

[Hal09] M. W. Hall. Comment on 'non-realism: deep thought or soft option?', by n. gisin, 2009.

[Hal10] Michael J. W. Hall. Local deterministic model of singlet state correlations based on relaxing measurement independence. *Phys. Rev. Lett.*, 105(25):250404–, December 2010.

[HBB99] Mark Hillery, Vladimír Bužek, and André Berthiaume. Quantum secret sharing. *Phys. Rev. A*, 59(3):1829–1834, March 1999.

[HGM+12] Martin Hendrych, Rodrigo Gallego, Michal Micuda, Nicolas Brunner, Antonio Acin, and Juan P. Torres. Experimental estimation of the dimension of classical and quantum systems. *Nat Phys*, 8(8):588–591, August 2012.

[HHH+10] K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, M. Pawlowski, and M. Bourennane. Contextuality offers device-independent security, 2010.

[HHHH09] Ryszard Horodecki, PaweÅ, Horodecki, MichaÅ, Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81(2):865–942, June 2009.

[Hol73] A. S. Holevo. *Prob. Peredachi Inf.*, 9, 3, 1973.

[HR83] Peter Heywood and Michael Redhead. Nonlocality and the kochen-specker paradox, 1983.

[JPPG+10] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf. Operator space theory: A natural framework for bell inequalities. *Phys. Rev. Lett.*, 104(17):170405–, April 2010.

[KHS+12] D. E. Koh, M. J. W. Hall, Setiawan, J. E. Pope, C. Marletto, A. Kay, V. Scarani, and A. Ekert. The effects of reduced "free will" on bell-based randomness expansion, 2012.

[Kly93] D.N. Klyshko. The bell and ghz theorems: a possible three-photon interference experiment and the question of nonlocality. *Physics Letters A*, 172(6):399–403, January 1993.

[KPB06] Johannes Kofler, Tomasz Paterek, and ÄŒaslav Brukner. Experimenter'ss freedom in bell's theorem and quantum cryptography. *Phys. Rev. A*, 73(2):022104–, February 2006.

[KS67] S. Kochen and E. P. Specker. *J. Math. Mech.*, 17:59, 1967.

[KSW+05] Nikolai Kiesel, Christian Schmid, Ulrich Weber, Géza Tóth, Otfried Gühne, Rupert Ursin, and Harald Weinfurter. Experimental analysis of a four-qubit photon cluster state. *Phys. Rev. Lett.*, 95(21):210502–, November 2005.

[Lap40] P. S. Laplace. *A phylosophical essay on probabilities.* 1840.

[LVB11] Yeong-Cherng Liang, Tamás Vértesi, and Nicolas Brunner. Semi-device-independent bounds on entanglement. *Phys. Rev. A*, 83(2):022108–, February 2011.

[MAG06] Ll. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73(1):012112–, January 2006.

[Mas02] L. Masanes. Tight bell inequality for d-outcome measurements correlations, 2002.

[Mas05] Ll. Masanes. Extremal quantum correlations for n parties with two dichotomic observables per site, 2005.

[Mas09] Lluis Masanes. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.*, 102(14):140501–, April 2009.

[Mer90a] N. David Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.*, 65(15):1838–1840, October 1990.

[Mer90b] N. David Mermin. Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.*, 65(27):3373–3376, December 1990.

[MTTT01] Gabriel Molina-Terriza, Juan P. Torres, and Lluis Torner. Management of the angular momentum of light: Preparation of photons in multidimensional vector states of angular momentum. *Phys. Rev. Lett.*, 88(1):013601–, December 2001.

[MTTT07] Gabriel Molina-Terriza, Juan P. Torres, and Lluis Torner. Twisted photons. *Nat. Phys.*, 3:305–310, 2007.

[MVWZ01] Alois Mair, Alipasha Vaziri, Gregor Weihs, and Anton Zeilinger. Entanglement of the orbital angular momentum states of photons. *Nature*, 412(6844):313–316, July 2001.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013–, 2008.

[PAM+10] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by bell's theorem. *Nature*, 464(7291):1021–1024, April 2010.

[PBS11a] Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. Extremal correlations of the tripartite no-signaling polytope. *Journal of Physics A: Mathematical and Theoretical*, 44(6):065303–, 2011.

[PBS+11b] E. Pomarico, J-D. Bancal, B. Sanguinetti, A. Rochdi, and N. Gisin. Various quantum nonlocality tests with a simple 2-photon entanglement source, 2011.

[Per91] A Peres. Two simple proofs of the kochen-specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175–, 1991.

[PGWP+08] D. Pérez-García, M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge. Unbounded violation of tripartite bell inequalities, 2008.

[PM11] S. Pironio and S. Massar. Security of practical private randomness generation, 2011.

[PPK+09] Marcin Pawlowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Zukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, October 2009.

[PR94] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, March 1994.

[PV08] Károly F. Pál and Tamás Vértesi. Efficiency of higher-dimensional hilbert spaces for the violation of bell inequalities. *Phys. Rev. A*, 77(4):042105–, April 2008.

[PZ10] Marcin Pawlowski and Marek Zukowski. Entanglement-assisted random access codes. *Phys. Rev. A*, 81(4):042326–, April 2010.

[RVC+09] Alessandro Rossi, Giuseppe Vallone, Andrea Chiuri, Francesco De Martini, and Paolo Mataloni. Multipath entanglement of two photons. *Phys. Rev. Lett.*, 102(15):153902–, April 2009.

[SASA05] Valerio Scarani, Antonio Acín, Emmanuel Schenck, and Markus Aspelmeyer. Nonlocality of cluster states of qubits. *Phys. Rev. A*, 71(4):042325–, April 2005.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *J. Comput. Syst. Sci.*, 33(1):75–87, 1986.

[Sve87] George Svetlichny. Distinguishing three-body from two-body nonseparability by a bell-type inequality. *Phys. Rev. D*, 35(10):3066–3069, May 1987.

[Tsi83] B. Tsirelson. *Lett. Math. Phys.*, 4:93, 1983.

[VDDMM09] Giuseppe Vallone, Gaia Donati, Francesco De Martini, and Paolo Mataloni. Polarization entanglement with graded-index lenses. *Appl. Phys. Lett.*, 95(18):181110–3, November 2009.

[VP08] T. Vértesi and K. F. Pál. Generalized clauser-horne-shimony-holt inequalities maximally violated by higher-dimensional systems. *Phys. Rev. A*, 77(4):042106–, April 2008.

[VP09] Tamás Vértesi and Károly F. Pál. Bounding the dimension of bipartite quantum systems. *Phys. Rev. A*, 79(4):042106–, April 2009.

[VPB10] Tamás Vértesi, Stefano Pironio, and Nicolas Brunner. Closing the detection loophole in bell experiments using qudits. *Phys. Rev. Lett.*, 104(6):060401–, February 2010.

[VPDMM08] Giuseppe Vallone, Enrico Pomarico, Francesco De Martini, and Paolo Mataloni. Active one-way quantum computation with two-photon four-qubit cluster states. *Phys. Rev. Lett.*, 100(16):160502–, April 2008.

[VV12] U. Vazirani and T. Vidick. Certifiable quantum dice: or, true random number generation secure against quantum adversaries. *Proceedings of the ACM Symposium on the Theory of Computing*, 2012.

[WCD08] Stephanie Wehner, Matthias Christandl, and Andrew C. Doherty. Lower bound on the dimension of a quantum system given measured data. *Phys. Rev. A*, 78(6):062112–, December 2008.

[WJS+98] Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger. Violation of bell's inequality under strict einstein locality conditions. *Phys. Rev. Lett.*, 81(23):5039–5043, December 1998.

[WPG09] Michael M. Wolf and David Perez-Garcia. Assessing quantum dimensionality from observable dynamics. *Phys. Rev. Lett.*, 102(19):190504–, May 2009.

[YCA+12] Tzyh Haur Yang, Daniel Cavalcanti, Mafalda L Almeida, Colin Teo, and Valerio Scarani. Information-causality and extremal tripartite correlations. *New Journal of Physics*, 14(1):013061–, 2012.

[YZZ+05] Tao Yang, Qiang Zhang, Jun Zhang, Juan Yin, Zhi Zhao, Marek Żukowski, Zeng-Bing Chen, and Jian-Wei Pan. All-versus-nothing violation of local realism by two-photon, four-dimensional entanglement. *Phys. Rev. Lett.*, 95(24):240406–, December 2005.

[ZG08] Stefan Zohren and Richard D. Gill. Maximal violation of the collins-gisin-linden-massar-popescu inequality for infinite dimensional states. *Phys. Rev. Lett.*, 100(12):120406–, March 2008.

[ZYC+03] Zhi Zhao, Tao Yang, Yu-Ao Chen, An-Ning Zhang, Marek Żukowski, and Jian-Wei Pan. Experimental violation of local realism by four-photon greenberger-horne-zeilinger entanglement. *Phys. Rev. Lett.*, 91(18):180401–, October 2003.