Programa de Profesores Visitantes
Fac. de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Nonlocality and Contextuality: Foundations and applications

Ana Belén Sainz

October 18, 2017

Introduction

La nolocalidad es uno de los aspectos más fundamentales y contraintuitivos de la física cuántica. El fenómeno demuestra que ciertas suposiciones sobre la Naturaleza, que parecen obvias y naturales (usualmente denominadas "localidad" y "realismo"), son incompatibles con la mecánica cuántica. La teoría formal que explica el fenómeno fue iniciada de manera rigurosa por J. Bell en 1964, y desde entonces varios experimentos lo han corroborado, siendo el 2015 el año en que los resultados definitivos fueron publicados. Más allá de su relevancia fundamental y filosófica, la nolocalidad juega un rol fundamental en la información cuántica, más precisamente en protocolos de tareas "device independent". En otras palabras, la nolocalidad es el recurso que permite hacer "distribución cuántica de claves" y "amplificación de aleatoriedad", entre otras, de manera segura en escenarios criptográficos.

Contextualidad es un fenómeno similar a la nolocalidad, en el sentido que desafía nuestra intuición clásica sobre los resultados obtenibles al hacer mediciones conjuntas en un sistema físico. Este fenómeno fue descubierto por Kochen y Specker en 1967, y desde entonces diferentes aspectos del mismo han sido desarrollados en varios formalismos. El estudio de las posibles aplicaciones de la contextualidad no fue profundizado sino hasta recientemente, cuando se descubrió que es un recurso necesario para desarrollar computaciones cuánticas.

Este curso está destinado a estudiantes del doctorado en física o del último año de la licenciatura. Nociones básicas de mecánica cuántica son requeridas, aunque en la primera unidad los conceptos necesarios serán revisados. El objetivo del curso es presentar los fenómenos de nolocalidad y contextualidad, sus aplicaciones y conceptos básicos sobre la historia experimental. Parte del curso se destinará a explorar estos fenómenos "mas allá de la mecánica cuántica", es decir, sin asumir que la física cuántica es la teoría que describe a la Naturaleza. A continuación, presento un temario tentativo por clase, de approximadamente tres horas cada una.

La literatura para este curso:

- Quantum Computation and Quantum Information, Michael A. Nielsen and Isaac L. Chuang, Cambridge University Press; 1 edition (2000).
- Lecture Notes for Physics 229: Quantum Information and Computation.

John Preskill, Caltech, Set 1998.
http://www.theory.caltech.edu/people/preskill/ph219/

- Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, Stephanie Wehner. Bell nonlocality. Rev. Mod. Phys. 86: 419, 2014.

El material más específico será revisado de artículos científicos.

Contents

1 Bell nonlocality – correlations

En esta unidad se presentan los escenarios de Bell y las condiciones que los sistemas clásicos satisfacen. Las hipótesis de Bell son presentadas en terminos de "causalidad local", y se derivan las desigualdades de Bell. El ejemplo que se trabaja en esta unidad es el escenario de Clauser-Horne-Shimony-Holt (CHSH). Luego se introducen (recuerdan) algunas nociones básicas de mecánica cuántica, como espacios de Hilbert, estados puros/mezcla/entrelazados, y mediciones. Las correlaciones cuánticas son presentadas (junto al dilation theorem), y así ejemplos de violaciones de desigualdades de Bell. Al final se comenta sobre diferentes nociones de correlaciones cuánticas y se menciona el problema de Tsirelson.

1.1 Bell experiment

Bell's seminal paper [1] was inspired by the exchange between Einstein-Podolsky-Rosen (EPR) [2] and Bohr. This EPR paradox, in a nutshell, poses a gedanken-experiment where entangled states display properties incompatible with a local and realistic world. The EPR paradox triggered a still ongoing philosophical debate on how to understand these nonclassical phenomena, and on whether Quantum theory is the ultimate and complete description of the world (for some definition of 'complete').

Nevertheless, Bell's theorem is not about quantum mechanics. Rather, it proves in a way independent of any specific physical theory, that the correlations among distant events cannot be arbitrarily strong if one assumes the validity of the principle of Local causality.

John Bell's breakthrough idea was to test 'Einstein locality' by considering the quantitative properties of the correlations between measurement outcomes obtained by two distant parties, Alice and Bob, who share a state. This kind of experiment may seem naïve or just *too simple*, but actually it is there that the discrepancy between classical physics and Nature manifests itself in the most straightforward way.

A Bell scenario hence consists of two parties, Alice and Bob, who can choose among m measurements to perform in their share of the system, each of which has d possible outcomes. In general one can assume that the number of measurements that Alice has access to is different to that of Bob, but here for simplicity we will consider them to be the same. Similarly, the measurements need not have the same number of outcomes, however we can consider them to be the same (and equal to that of the measurement with the largest number of outcomes) by formally assigning a probability zero to the outcomes beyond those that each particular measurement has. Finally, one can also consider Bell scenarios with more than two parties. This is going to be discussed in a a later Lecture, and now we will just focus on bipartite Bell scenarios.

In a Bell experiment then, Alice and Bob have access to a large number of independent copies of system, and in each round of the experiment they:

- take a new copy of the system.
- choose randomly and independently which measurement to perform. Alice's choice is usually denoted by $x \in \{1, \ldots, m\}$, and Bob's by $y \in \{1, \ldots, m\}$. tem[-] perform the measurements and obtain the outcomes a for Alice and b for Bob.
- keep track of the *event* $(ab|xy)$.

After performing a large number of rounds, Alice and Bob get together and compute the conditional probability distribution $p(ab|xy)$. This p is also usually referred to as *correlations* or *behaviour*.

The aim of Bell's theorem then is to draw conclusions on the underlying theory that generates those p. More precisely, Bell's theorem characterises those correlations that may have an explanation coming from a physical theory that obeys local causality.

Note that the only information that is used in this experiment is the classical labels of the measurement choices and the classical label of their outputs. No information on the inner working of the measurement devices is required, which may actually not be performing the measurements we think x and y relate to. The measurement apparatuses are hence thought of as black boxes, as depicted in Fig. 1.1. Hence, the study of correlations in a Bell scenario this way is also said to be in a *device independent framework*, which is of utmost relevance in cryptographic tasks.

Finally, a Bell experiment further has some implicit assumptions that include the following [3]:

-Space-Time: The concepts of space-like separation, light-cones, etc. can be applied unambiguously in ordinary laboratory situations.

-Arrow of time: A cause can only be in the past of its effect.

-Free choice: x and y are freely chosen, and hence independent of the past and independent of each other.

-Relativistic Causality: In relation to causation, 'the past' is to be understood as 'the past light cone'.

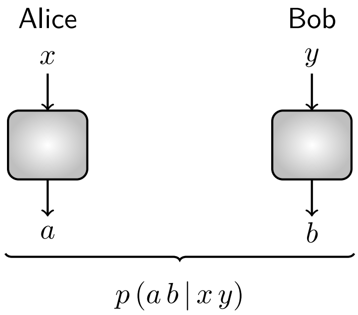


Figure 1.1: Bell experiment: Alice inputs her measurement choice x on her measurement apparatus, depicted by a black box, and obtains an outcome a. Bob similarly inputs y and obtains b. By performing this many times on identical and independent copies of the shared system, Alice and Bob can compute the conditional probability distribution $p(ab|xy)$.

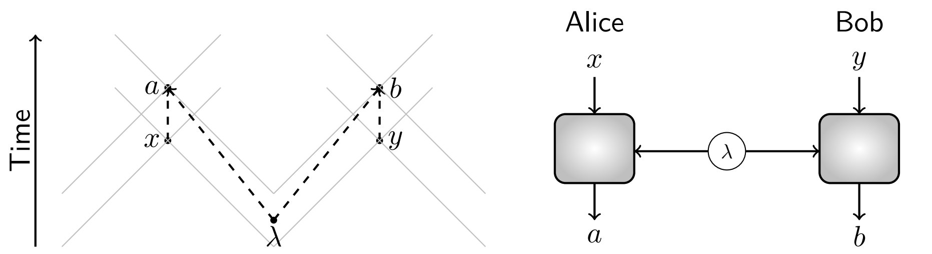
In addition, we will restrict ourselves to Bell experiments where the time-window of Alice's 'measurement' (i.e. when chooses her measurement, implements it and obtains the outcome) is outside the future light cone of Bob's 'measurement', and vice-versa. That is, we will assume that for each measurement round Alice and Bob's actions are space-like separated.

1.2 Locally causal models

As we mentioned before, Bell's theorem characterises the correlations that may be obtained by performing measurements on a system that is governed by the laws of a physical theory that is locally causal (LC). Historically, these correlations are said to have a *local hidden variable* (LHV) model.

Classical mechanics is an example of a theory that satisfies LC. But indeed, Bell's first version of his theorem didn't actually rely on the notion of LC [4] but rather on the conjunction of the concepts of *locality* and *determinism* (LD) [1].

While LC is a strictly weaker concept than LD [3], it was highlighted by Fine in 1982 [5] that the range of phenomena respecting LD is the same as the range of phenomena respecting LC. Hence, these two may be thought of as different interpretations of the Bell's theorem, each of which allows to draw a particular conclusion on the properties that Nature should not respect. On the one hand, should there exists correlations without a LD model, one should accept that physical phenomena violate either determinism, or locality, or both. On the other hand, if



(a) Space-time diagram of a locally causal model (b) A locally causal model, operationally

Figure 1.2: Representation of a locally causal model. (a) The space-time diagram depicting the light-cones of each classical variable. The classial common source cannot influence the choice of measurements x and y. Even taking into account the duration of the measurement process, the choice of x in Alice's lab cannot influence Bob's outcome b, and vice-versa. (ii) Black-box diagram, where a source prepares the shared randomness λ, which is distributed between Alice and Bob. After inputting the settings x (y), together with the variable λ ab outcome a (b) is output from the black-box, which represents the measurment device.

we think of these correlations instead as not having a LC model one must accept that physical phenomena violate LC. Whether LD or LC is the correct philosophical way to interpret a classical theory is beyond the scope of these lectures, and for personal preference the case of LC models will be presented.

Let us denote by λ a classical random variable representing the common cause to Alice and Bob's actions. This λ contains the relevant information that appears in the common past of both Alice and Bob. A behaviour p hence has a LC model if the correlations between a and b can be accounted for via λ. We represent this situation schematically in Fig. 1.2: the common cause λ influences the local response functions $p_A(a|x,\lambda)$ and $p_B(b|y,\lambda)$ with which the measurement devices produce the outcomes a and b. The effective correlations hence are those that arise after averaging over the random variable λ, that is

$$p_{LC}(ab|xy) = \int d\lambda\, q(\lambda)\, p_A(a|x,\lambda)\, p_B(b|y,\lambda)\,, \tag{1.1}$$

where $q(\lambda)$ is the distribution over the random variable λ. The local response functions $p_A(a|x,\lambda)$ and $p_B(b|y,\lambda)$ are moreover normalised conditional probability distributions for each λ.

1.3 Locally causal models and Fine's theorem

Even though the notion of Local causality may appeal to some physicists, truth is, when hands-on time comes it is easier to handle local deterministic models. We will hence review Fine's argument that a behaviour has a LC model iff it has a LD one [5].

First, correlations that have a LD model are those which can be written as

$$p_{LD}(ab|xy) = \int d\lambda\, q(\lambda)\, D_A(a|x,\lambda)\, D_B(b|y,\lambda)\,, \tag{1.2}$$

where $q(\lambda)$ is the distribution over the random variable λ. The local response functions $D_A(a|x,\lambda)$ and $D_B(b|y,\lambda)$ are not just normalised but also deterministic conditional probability distributions for each λ, i.e. $D_A(a|x,\lambda) \in \{0,1\}$ and $D_B(b|y,\lambda) \in \{0,1\}$.

Hence, if a behaviour has a model as in eq. (1.2), it immediately has a locally causal one as in eq. (1.1).

To see that the converse also holds, first notice that every probability distribution can be decomposed into a convex combination of deterministic probabilities. Hence, $p_A(a|x,\lambda) = \int d\mu q(\mu)\, D_A(a|x,\lambda,\mu)$, and similarly $p_B(b|y,\lambda) = \int d\nu q(\nu)\, D_B(b|y,\lambda,\nu)$. Hence, a behaviour p with a locally causal may be expressed as

$$p_{LC}(ab|xy) = \int d\lambda d\mu d\nu\, q(\lambda)q(\mu)q(\nu)\, D_A(a|x,\lambda,\mu)\, D_B(b|y,\lambda,\nu)\,.$$

Now by redefining the hidden variables as $\tilde{\lambda} := (\lambda,\mu,\nu)$, we can rewrite it as

$$p_{LC}(ab|xy) = \int d\tilde{\lambda}\, q(\tilde{\lambda})\, D_A(a|x,\tilde{\lambda})\, D_B(b|y,\tilde{\lambda})\,,$$

where $q(\tilde{\lambda}) := q(\lambda)q(\mu)q(\nu)$ is a distribution over this new random variable $\tilde{\lambda}$. Note that the expression to which we have arrived is a LD model for p_{LC}. We see hence that if a behaviour has a LC model then it also has a LD one.

1.4 Bell's theorem and Bell inequalities

The line of reasoning behind Bell's theorem is the following:

1- Find a functional I (a.k.a. Bell expression) on the conditional probability distributions $p(ab|xy)$.

2- Compute the maximum value β_{LC} that $I\,(p(ab|xy))$ can take when the correlations have a LC model as in eq. (1.1).

3- Show that there exist quantum correlations that yield a value of $I\left(p(ab|xy)\right)$ larger than β_{LC}.

The beauty of the argument is its simplicity which nevertheless has such a great power. Back then, Bell's theorem provided not just a novel idea, but also a suitable functional I whose search required craftsmanship. Since then, every functional I whose maximum value for LC models is bounded, together with its β_{LC}, are referred to as a 'Bell Inequality'. After Bell's paper, many effort was devoted to the search of new relevant Bell inequalities for different Bell scenarios, and focused mainly on linear Bell expressions. The study also shifted to the search for quantum informationally relevant inequalities, or Bell functionals that were experimentally friendlier. As of today, Bell inequalities are thought of as a complex Zoo.

Now we will present the derivation of the most famous Bell inequality, the one derived by Clauser, Horne, Shimony and Holt and therefore known as CHSH [6]. Consider then a Bell scenario where Alice and Bob can each choose from between two dichotomic measurements. The measurements are labelled by $\{0, 1\}$, and their outcomes as well. The functional I that CHSH proposed is the following:

$$I\left(p\right) = \left|E_{00} + E_{01} + E_{10} - E_{11}\right| \,, \tag{1.3}$$

where E_{xy} are the correlators defined as

$$E_{xy} = p(00|xy) + p(11|xy) - p(01|xy) - p(10|xy)\,.$$

The challenge now is to compute the maximum value of eq. (1.3) for LC models. The first step is to notice that, for each λ the correlators take the form

$$E^{\lambda}_{xy} = E^{\lambda}_{x}\, E^{\lambda}_{y}\,,$$

where $E^{\lambda}_{x} = p_A(1|x,\lambda) - p_A(0|x,\lambda)$, and similarly for Bob. Hence,

$$\begin{aligned} I\left(p_{LC}\right) &= \left|\int d\lambda\, q(\lambda)\left(E^{\lambda}_{x=0}\left(E^{\lambda}_{y=0} + E^{\lambda}_{y=1}\right) - E^{\lambda}_{x=1}\left(E^{\lambda}_{y=0} - E^{\lambda}_{y=1}\right)\right)\right| \\ &\leq \int d\lambda\, q(\lambda)\left|\left(E^{\lambda}_{x=0}\left(E^{\lambda}_{y=0} + E^{\lambda}_{y=1}\right) - E^{\lambda}_{x=1}\left(E^{\lambda}_{y=0} - E^{\lambda}_{y=1}\right)\right)\right|\,. \end{aligned} \tag{1.4}$$

To compute the maximum value of eq. (1.4) we can assume with no loss of generality that the local response functions are deterministic, and hence $E^{\lambda}_{x} = \pm 1$ and $E^{\lambda}_{y} = \pm 1$. Direct inspection then shows that

$$\left|\left(E^{\lambda}_{x=0}\left(E^{\lambda}_{y=0} + E^{\lambda}_{y=1}\right) - E^{\lambda}_{x=1}\left(E^{\lambda}_{y=0} - E^{\lambda}_{y=1}\right)\right)\right| \leq 2 \quad \forall\, \lambda\,,$$

and hence the CHSH Bell inequality reads

$$I\left(p\right) = \left|E_{00} + E_{01} + E_{10} - E_{11}\right| \leq 2\,. \tag{1.5}$$

The bound in eq. (1.5) can moreover be saturated. Consider for instance the behaviour $p^*(ab|xy) = \delta_{a,0}\,\delta_{b,0}$, i.e. both Alice and Bob always output 0 independently of their measurement settings. This conditional probability distribution is factorizable and deterministic, hence it is written as a LD model of eq. (1.2). For this behaviour, $E_{x=0} = -1$, $E_{x=1} = -1$, $E_{y=0} = -1$ and $E_{y=1} = -1$. Hence, eq. (1.3) achieves a value of 2 and the CHSH inequality of (1.5) is saturated.

Next we will see how quantum mechanics, and also Nature, violate the CHSH inequality. But before, a remark on tightness is in order. In the literature, people talk about "tight Bell inequalities", but sometimes they mean different things. On the one hand, some people denote as 'tight' Bell inequalities those which are saturated by some LC behaviours. On the other hand, some denote as 'tight' Bell inequalities those that are saturated by at least a certain number of LD behaviours, where the quantity depends on the Bell scenario (number of parties, measurements and outcomes) under study. The second notion, which is stronger than the former, will be formalised later on when studying the *Bell polytope*.

1.5 Quantum correlations

In order to present the correlations that are allowed by quantum mechanics, we will first review some basics concepts regarding quantum theory.

In quantum theory, the state of a system is an element of a Hilbert space $\mathcal{H}$, represented by a positive semidefinite matrix ρ, usually called *density matrix*. A special class of states is that of *pure* quantum states, which correspond to vectors $|\Psi\rangle$ over the Hilbert space. In this case, the density matrix is given by $\rho = |\Psi\rangle\langle\Psi|$. The observables, moreover, are self-adjoint operators $\mathcal{A}$ on $\mathcal{H}$, whose expectation values are given by the Born's rule $\langle\mathcal{A}\rangle = \mathrm{tr}\{\mathcal{A}\,\rho\}$. The most general class of measurements over quantum systems is called *positive operator-valued measure* (POVM) [7]. There, a measurement x is described by a set of nonnegative operators $\{M_a^x\}$ with the following properties:

- $\sum_a M_a^x = \mathbb{1}_{\mathcal{H}}$,
- each operator M_a^x is associated to a possible outcome of the measurement, so that the probability of obtaining a when measuring x is given by $p(a|x) = \mathrm{tr}\{\rho\, M_a^x\}$.

The nonnegativity of $\{M_a^x\}$ assures that the $p(a|x)$ are positive numbers, and the condition that $\{M_a^x\}$ sum up to the identity guarantees the probabilities $p(a|x)$ to be normalized. Note that these operators M_a^x need not be projectors over $\mathcal{H}$. When they are, the measurement belongs to a smaller family called *projective* or *von Neumann measurements*.

An interesting property is that, given a general state ρ and POVM $\{M_a^x\}$ in a Hilbert space $\mathcal{H}$, it is always possible to find a Hilbert space $\mathcal{H}'$ of larger dimension, a state ρ' and a projective measurement $\{\Pi_a^x\}$, with identical statistics for the measurement outputs, i.e. $p(a|x) = \mathrm{tr}\,\{\rho\, M_a^x\} = \mathrm{tr}\,\{\rho'\,\Pi_a^x\}$ [7]. Indeed, suppose that we want to perform a measurement $\{M_a^x\}$ on the system ρ. Consider an ancillary system belonging to a Hilbert space $\mathcal{H}_b$, such that there exists a basis of orthonormal states $\{|a\rangle\}$ in $\mathcal{H}_b$ in one-to-one correspondence with the measurement outcomes of $\{M_a^x\}$ over ρ. This ancillary system can be thought of as a purely mathematical device appearing in the construction, or as an actual quantum physical system that helps in the measurement process. Operationally, the main idea then is to perform an entangling operation between the system ρ and the ancilla, which contains the information about the original POVM, and then perform a projective measurement $\{\mathbb{1}_\mathcal{H} \otimes |a\rangle\langle a|\}$ over the state of the ancilla. Formally, the construction of $\{\Pi_a^x\}$ from $\{M_a^x\}$ goes as follows. Since $\{M_a^x\}$ are positive semidefinite operators, they may be expressed as[1] $M_a^x = K_a^{x\dagger} K_a^x$. Consider now the initial joint state $\rho' = \rho \otimes |0\rangle\langle 0|$, where $|0\rangle\langle 0|$ is the state of the ancilla, and define the unitary U which performs the entangling operation $U\rho' U^\dagger = \sum_{a,a'} K_a^x \rho K_{a'}^{x\dagger} |a\rangle\langle a'|$. Finally, the operators $\Pi_a^x = U^\dagger(\mathbb{1}_\mathcal{H} \otimes |a\rangle\langle a|)U$ indeed form a projective measurement over $\mathcal{H}' = \mathcal{H}\otimes\mathcal{H}_b$, with $\mathrm{tr}\,\{\Pi_a^x \rho'\} = p(a|x)$.

The next step towards defining quantum correlations involves the notion of composite system. Now the simplest composite scenario will be presented, and the general discussion will be left for later (Tsirelson's problem).

Let us consider the case of two parties, Alice and Bob, whose local Hilbert spaces are $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The Hilbert space $\mathcal{H}$ describing the joint situation is defined as the tensor product of the individial Hilbert spaces, i.e. $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Let ρ be a quantum state in $\mathcal{H}$ shared by Alice and Bob. Let $\{M_a^x\}$ define a POVM x for Alice, for each $x \in \{0, \ldots, m-1\}$, and similarly for Bob. Hence, Born's rule tells that the correlations between Alice's and Bob's measurement outcomes are given by:

$$p(ab|xy) = \mathrm{tr}\,\{M_a^x \otimes M_b^y\,\rho\}\ . \tag{1.6}$$

1.6 Nonlocal quantum correlations

Quatum theory may exhibit Bell nonlocality, i.e. there exist quantum correlations that violate Bell inequalities. Now we will see an example of a state and measurements that generate statistics that violate the CHSH inequality (1.5).

[1] The operators K_a^x are usually called *Kraus operators*. The decomposition of the elements of a POVM into its Kraus operators is not unique, since any unitaries acting on $\{K_a^x\}$ preserves the form of $\{M_a^x\}$.

Let Alice and Bob share the singlet state $|\Psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$. Take as Alice's measurements

$$M_a^{x=0} = \frac{\mathbb{1} + (-1)^a \, \sigma_x}{2} \, ,$$
$$M_a^{x=1} = \frac{\mathbb{1} + (-1)^a \, \sigma_z}{2} \, ,$$

where $\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is the Pauli-x matrix, and $\sigma_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ is the Pauli-z. As Bob's measurements define

$$M_b^{y=0} = \frac{\mathbb{1} + \frac{(-1)^b}{\sqrt{2}} \, (\sigma_x + \sigma_z)}{2} \, ,$$
$$M_b^{y=1} = \frac{\mathbb{1} + \frac{(-1)^b}{\sqrt{2}} \, (\sigma_x - \sigma_z)}{2} \, .$$

On the one hand, note that

$$\begin{aligned} E_{xy} &= \langle M_0^x \otimes M_0^y - M_0^x \otimes M_1^y - M_1^x \otimes M_0^y + M_1^x \otimes M_1^y \rangle_\rho \\ &= \langle (M_0^x - M_1^x) \otimes (M_0^y - M_1^y) \rangle_\rho \, . \end{aligned}$$

On the other, note that

$$M_0^x - M_1^x = \vec{x} \cdot \sigma \, ,$$

where $\vec{x} \cdot \sigma = x_x \sigma_x + x_y \sigma_y + x_z \sigma_z$. Which makes

$$E_{xy} = \langle (\vec{x} \cdot \sigma) \otimes (\vec{y} \cdot \sigma) \rangle_\rho = -\vec{x} \cdot \vec{y} \, .$$

In this notation, this example has $x = 0 \Rightarrow \vec{x} = (1,0,0)$, $x = 1 \Rightarrow \vec{x} = (0,0,1)$, $y = 0 \Rightarrow \vec{y} = (1,0,1)/\sqrt{2}$ and $y = 1 \Rightarrow \vec{y} = (1,0,-1)/\sqrt{2}$. Hence, the correlations these state and measurements define yield $I(p) = 2\sqrt{2} > 2$. This shows that there exist quantum correlations that go beyond what is admissible within a locally causal theory.

John Preskill, in his notes [8], gives the following comment on this counter-intuitive aspect of quantum theory:

> *"The human mind seems to be poorly equipped to grasp the correlations exhibited by entangled quantum states, and so we speak of the weirdness of quantum theory. But whatever your attitude, experiment forces you to accept the existence of the weird correlations among the*

measurement outcomes. There is no big mystery about how the correlations were established – we saw that it was necessary for Alice and Bob to get together at some point to create entanglement among their qubits. The novelty is that, even when A and B are distantly separated, we cannot accurately regard A and B as two separate qubits, and use classical information to characterize how they are correlated. They are more than just correlated, they are a single inseparable entity, they are entangled."

1.7 The set of quantum correlations

So far we studied correlations that are compatible with a locally causal explanation. Now we will focus on those correlations that are admissible within quantum theory. We will assume in the rest of the course that the dimensions of the Hilbert spaces are all finite, and hence we will not worry about Tsirelson's problem (see next section).

A conditional probability distribution $p(ab|xy)$ is said to have a quantum explanation (or quantum realisation) if there exists a Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, a projective measurement $\{\Pi_{a|x}\}_a$ for each x for Alice, a projective measurement $\{\Pi_{b|y}\}_b$ for each y for Bob, and a quantum state ρ in $\mathcal{H}$ such that the statistics are recovered by them, i.e.

$$p(ab|xy) = \mathrm{tr}\left\{\Pi_{a|x} \otimes \Pi_{b|y}\, \rho\right\} .$$

The set of correlations that have such a quantum realisation is called the *quantum set* and is usually denoted by $\mathcal{Q}$.

Notice that in the definition we have restricted ourselved to measurements for Alice and Bob that are projective. A valid question is then whether relaxing that condition to allow for POVMs instead may allow within the quantum set correlations that otherwise wouldn't be explainable. This question is related to Problem **??** which can be answered in the positive. That is, a correlation realisable via POVMs can always be equivalently realised by projective measurements taking a Hilbert space of larger dimension. Hence, since there are no restrictions on the Hilbert space dimensions in the definition of the quantum set (i.e. we can take it as large as want while keeping it finite), POVM realisable correlations all belong to it.

This property of quantum correlations in Bell scenarios is usually referred to as *dilation*, and more colloquially as *the Church of the larger Hilbert space*. This dilation theorem has been proven in different instances by various mathematical techniques. the most popular one given by Naimark [9], and can be thought of as a consequence of Stinespring's dilation theorem. In what follows we will review a dilation proof by Vern Paulsen (Theorem 9.8 in [10]) that uses similar mathematical tools to those in these lectures. The reader interested in Operator Algebras can review the other equivalent dilation theorems: double-Stinespring theorem [11], Naimark's dilation theorem [9], or Chapter 4 in [12].

1.8 Comment on Tsirelson's problem

A brief comment is in order about what has been known by today as "Tsirelson's problem" [13]. This problem is related to the way of computing correlations in composite systems, which I skilfully omitted to discuss so far. For the sake of the argument it is enough to consider the case of two parties, who perform space-like separated actions in distant labs.

On the one hand, the "tensor product paradigm" tells us that the way to describe the situation is by assigning to Alice a Hilbert space $\mathcal{H}_A$ and to Bob one $\mathcal{H}_B$, and defining the joint Hilbert space as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Then, Alice's measurement operators will be POVMs in $\mathcal{H}_A$ and Bob's POVMs in $\mathcal{H}_B$, while the shared quantum state ρ is a density matrix in $\mathcal{H}$ and the outcome probabilities are given by

$$p_T(ab|xy) = \mathrm{tr}\left\{M_a^x \otimes M_b^y \,\rho\right\} .$$

This is indeed what was presented in the previous sections.

However, there is another paradigm called "commutativity paradigm" that describes instead the situation as follows. Alice and Bob are both described by the same joint Hilbert space $\mathcal{H}$, and Alice's as well as Bob's measurements are POVMs in $\mathcal{H}$. The fact that Alice and Bob operate in a space-like separated manner is taken into account by imposing that the POVMs on her side commute with those of his. That is, $[M_a^x, M_b^y] = 0$ for all a, b, x, y. The correlations in this paradigm are defined as

$$p_C(ab|xy) = \mathrm{tr}\left\{M_a^x M_b^y \,\rho\right\} .$$

Correlations admissable by the tensor product paradigm can always be understood within the commutativity one by thinking of Alice's measurement operators in the joint Hilbert space as $M_a^x := M_a^x \otimes \mathbb{1}_B$, and similarly for Bob, since the tensor product guarantees that these new measurements do commute.

Conversely, whenever the dimension of $\mathcal{H}$ is finite, it has been proven that correlations admissible by the commutatibity paradigm can be understood within the tensor product one. A proof of this can be found in [14] (see also [13]) and will not be discussed in this lecture due to its complexity.

However, when the dimension of $\mathcal{H}$ is infinite the situation changes: there exist correlations in the commutativity paradigm that cannot be explained by the tensor product one [15]. Again, the proof of this statement gies beyond the scope of these lectures, but anyone who is not taken back by complex maths is welcome to read the paper.

The physically relevant question now is whether there exists an experiment that can settle this dispute: i.e. if we can measure correlations $p_C(ab|xy)$ that do not have a tensor product explanation, Nature will tell us that the correct way of describing composite systems is indeed the commutativity paradigm. Our main problem when attemtping this are experimental imperfections and error: we will never measure $p_C(ab|xy)$ with infinite accuracy. Now, should the set of correlations $p_T(ab|xy)$ have a completion equivalent to the set commutativity correlations, then this whole approach will be doomed, since there would always be a correlation in the tensor product paradigm that could explain $p_C(ab|xy)$ for any fixed precision. Whether these sets have indeed this type of equivalence is still an open question.

2 Quantum nonlocality

En esta unidad se presentan resultados sobre el fenómeno de nolocalidad en sistemas cuánticos. Se demuestra que no todo estado entrelazado exhibe correlaciones nolocales, pero que todo estado puro entrelazado es nolocal. Luego se discute la activación de la nolocalidad. Finalmente, se presenta una breve recopilación de experimentos realizados, desde los pioneros de Aspect en los '80, hasta los definitos del 2015, haciendo hincapié en los tecnicismos (loopholes) que fueron superados.

2.1 Entanglement vs nonlocality

Classical physics, as being a locally causal theory, cannot exhibit violations of Bell inequalities. Quantum theory, on the the other hand, as we have seen may display nonlocal behaviours. Since entanglement is a key feature of quantum mechanics that has no classical analogue, a valid question is how it relates to the phenomenon of Bell nonlocality. As we will see in what follows, entanglement is indeed a necessary condition for nonlocality, although it not always proves sufficient.

Let us first briefly review the notion of entanglement. Entanglement arises when describing composite systems: a pure state that cannot be written as a product state is called entangled. For mixed states, on the other hand, the definition is more subtle. First, a mixed state ρ is called separable if it can be written as a convex combination of product states, i.e.

$$\rho = \sum_k c_k \rho_k^A \otimes \rho_k^B \, .$$

If such a decomposition exists, then the state could have been equivalently prepared locally by the parties via the following classical protocol: with probability c_k prepare ρ_k^A in Alice's lab and ρ_k^B in Bob's. A mixed state is said to be entangled then if it is not separable.

Now, how does entanglement relate to nonlocality? Let us consider first the case where Alice and Bob perform measurements on a normalised separable state ρ_{sep}. We will assume the measurements to be POVMs, not to loose generality[1].

[1]The fact that we need to consider POVMs does not contradict the before mentioned dilation

The correlations that may arise in such manner are the following:

$$\begin{aligned}p(ab|xy) &= \mathrm{tr}\left\{M_a^x \otimes M_b^y \, \rho_{\mathsf{sep}}\right\} \\ &= \sum_k c_k \,\mathrm{tr}\left\{M_a^x \otimes M_b^y \, \rho_k^A \otimes \rho_k^B\right\} \\ &= \sum_k c_k \, p_k^A(a|x)\, p_k^B(b|y)\,.\end{aligned}$$

That is, the correlations have a locally causal model. This means that whenever Alice and Bob share a separable state, regardless of the measurements they perform they will always be able to only generate LC conditional probability distributions.

We learn then that entanglement is a necessary condition for the state shared by Alice and Bob to be able to display nonlocality under suitable measurements. Now the question is whether entanglement is indeed sufficient to guarantee that the state can produce nonlocal correlations. The answer to this question depends strongly on the purity of the state as we will see below.

Pure states.–
First consider the case of pure entangled states. Here, it was shown [16] that any (bipartite) pure entangled state can violate a Bell inequality. The argument goes as follows. First, take the Schmidt decomposition of the entangled state $|\Psi\rangle$, i.e. $|\Psi\rangle = \sum_k \alpha_k \, |\phi\rangle_k \otimes |\varphi\rangle_k$. Since the state is entangled we know that at least $\alpha_1 \neq 0 \neq \alpha_2$. Define then $|\xi\rangle = \alpha_1 \, |\phi\rangle_1 \otimes |\varphi\rangle_1 + \alpha_2 \, |\phi\rangle_2 \otimes |\varphi\rangle_2$, and $\left|\xi^\perp\right\rangle = \sum_{k=3} \alpha_k \, |\phi\rangle_k \otimes |\varphi\rangle_k$. Note that $|\xi\rangle \perp \left|\xi^\perp\right\rangle$. Let us define the computational basis via the Schmidt one as $|\phi\rangle_1 \otimes |\varphi\rangle_1 = |00\rangle$ and $|\phi\rangle_2 \otimes |\varphi\rangle_2 = |11\rangle$.

The idea now is to show that $|\Psi\rangle$ can violate the CHSH inequality (1.5) for suitable choices of the measurement operators. The correlators E_{xy} in the inequality can indeed be computed by the expectation value of the corresponding dichotomic observables as we implicitly used before: $E_{xy} = \langle A_x B_y\rangle$. Hence, parametrize dichotomic observables for Alice and Bob as follows:

$$\begin{aligned}A(\theta) &:= \cos(\theta)\sigma_z + \sin(\theta)\sigma_x\,, \\ B(\vartheta) &:= \cos(\vartheta)\sigma_z + \sin(\vartheta)\sigma_x\,.\end{aligned}$$

The state $|\xi\rangle$ has the properties

$$\begin{aligned}\langle \sigma_x \otimes \sigma_x\rangle_{|\xi\rangle} = 2\,\alpha_1\,\alpha_2\,, \quad \langle \sigma_z \otimes \sigma_z\rangle_{|\xi\rangle} &= 1\,, \\ \langle \sigma_x \otimes \sigma_z\rangle_{|\xi\rangle} = 0 &= \langle \sigma_z \otimes \sigma_x\rangle_{|\xi\rangle}\,.\end{aligned}$$

theorems, since the two questions are different in nature. One asks given the correlations whether they may have a quantum realisation, and the other asks given a state which are the correlations that may arise from it.

Therefore, the correlators for a given choice of θ and ϑ are

$$\langle A(\theta)\, B(\vartheta)\rangle_{|\xi\rangle} = \cos(\theta)\, \cos(\vartheta) + 2\,\alpha_1\,\alpha_2\, \sin(\theta)\, \sin(\vartheta)\,.$$

Now take two specific dichotomic observables per party as follows: $A_0 = A(0)$, $A_1 = A(\frac{\pi}{2})$, $B_0 = B(\vartheta)$ and $B_1 = B(-\vartheta)$. The correlators these observables yield on $|\xi\rangle$ are the following:

$$\begin{aligned}\langle A_0\, B_0\rangle_{|\xi\rangle} &= \cos(\vartheta) = \langle A_0\, B_1\rangle_{|\xi\rangle}\,,\\ \langle A_1\, B_0\rangle_{|\xi\rangle} &= 2\,\alpha_1\,\alpha_2\, \sin(\vartheta) = -\langle A_1\, B_1\rangle_{|\xi\rangle}\,.\end{aligned}$$

The Bell functional for the CHSH inequality evaluated on these state and measurements hence yields:

$$I(p) = 2\,|\cos(\vartheta) + 2\,\alpha_1\,\alpha_2\, \sin(\vartheta)|\,. \tag{2.1}$$

Expanding to linear order in ϑ, we obtain

$$I(p) \sim 2\,|1 + 2\,\alpha_1\,\alpha_2\,\vartheta|\,, \tag{2.2}$$

which gives a value $I(p) > 2$ for $\alpha_1\,\alpha_2 > 0$ and ϑ positive and small.

The last part of the proof relies on noticing that these observables yield the same value for the CHSH Bell functional when measured on the state $|\Psi\rangle$, since their action on the orthogonal complement of the subspace defined by $|\xi\rangle$ is null.

Mixed states.–

For the case of mixed entangled state, the situation is not as elegant as for pure ones. Indeed, the claim for pure entangled states cannot be extended to mixed ones since there exist entangled states that may not violate any Bell inequality regardless of the choice of measurement settings. In what follows we will discuss the example of Werner states.

A two-qubit Werner state is defined as as convex combination of a maximally entangled state and a maximally mixed one, i.e.

$$\rho_r^W = r\,|\Psi\rangle\,\langle\Psi| + (1-r)\,\frac{\mathbb{1}}{4}\,, \tag{2.3}$$

where $|\Psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$ and $r \in [0,1]$. This state is entangled whenever $r > \frac{1}{3}$. Now, we will first show that the correlations that may arise when performing any number of projective measurements on ρ_r^W for $r = \frac{1}{2}$ have always a locally causal explanation. This was first shown by Werner [17], and here we give the explicit construction following the presentation of [18]. Define two arbitrary projective measurements for Alice and Bob as follows:

$$M_{\vec{x}}^{A} = \frac{\mathbb{1} + \vec{x}\cdot\sigma}{2}\,,$$
$$M_{\vec{y}}^{B} = \frac{\mathbb{1} + \vec{y}\cdot\sigma}{2}\,,$$

where we use the notation of Section 1.6. The normalised vectors $\vec{x}$ and $\vec{y}$ describe the measurements in the Bloch sphere, and also denote the 'direction in which the spin polarisation is measured' when the system of two levels probed in the experiment consists of photons with spins 'up' or 'down'. The correlations between the '0' outcomes of Alice and Bob when measuring ρ_r^W are given by

$$p_r(00|\vec{x}\vec{y}) = \frac{1}{4}\left(1 - r\,\vec{x}\cdot\vec{y}\right)\,.$$

In what follows, we present a local hidden variable model that gives the same statistics.

Let the hidden variable λ be a vector that denotes a direction in the Bloch sphere: $\lambda = (\sin\theta\,\cos\phi, \sin\theta\,\sin\phi, \cos\theta)$. This λ is known by both Alice and Bob, and in each round of the experiment a different λ is sent chosen according to the uniform distribution. The local response functions for Alice and Bob are defined as

$$p_A(0|\vec{x},\lambda) = \cos^2\left(\frac{\alpha_A}{2}\right)\,,$$

where $\cos(\alpha_A) = \vec{x}\cdot\lambda$, and

$$p_B(0|\vec{y},\lambda) = \begin{cases} 1 & \text{if } 2\cos^2\left(\frac{\alpha_B}{2}\right) < 1\,, \\ 0 & \text{if } 2\cos^2\left(\frac{\alpha_B}{2}\right) > 1\,, \end{cases}$$

where $\cos(\alpha_B) = \vec{y}\cdot\lambda$. Then, the bipartite correlations that this LHV model produces are given by

$$p_{\mathsf{LHV}}(00|\vec{x}\vec{y}) = \int p_A(0|\vec{x},\lambda)\, p_B(0|\vec{y},\lambda)\, \frac{\sin\theta}{4\pi} d\theta\, d\phi\,.$$

Problem 2.1. Show that $p_{\mathsf{LHV}}(00|\vec{x}\vec{y}) = p_r(00|\vec{x}\vec{y})$ for $r = \frac{1}{2}$.

Proof. The only terms that contribute to the integral are those where $2\cos^2\left(\frac{\alpha_B}{2}\right) < 1$, i.e. whenever $\vec{y}\cdot\lambda < 0$. Since we are integrating over the whole Bloch sphere, let us assume that the direction given my measurement $\vec{y}$ coincides with $(0,0,1)$. Hence,

$$
\begin{aligned}
p_{\mathsf{LHV}}(00|\vec{x}\vec{y}) &= \int_{\frac{\pi}{2}}^{\pi} d\theta \int_0^{2\pi} d\phi \, p_A(0|\vec{x},\lambda)\,\frac{\sin\theta}{4\pi}\,, \\
&= \int_{\frac{\pi}{2}}^{\pi} d\theta \int_0^{2\pi} d\phi \,\frac{\vec{x}\cdot\lambda+1}{2}\,\frac{\sin\theta}{4\pi}\,, \\
&= \frac{1}{8\pi}\int_{\frac{\pi}{2}}^{\pi}\sin\theta d\theta\int_0^{2\pi} d\phi + \frac{1}{8\pi}\int_{\frac{\pi}{2}}^{\pi}\sin\theta d\theta\int_0^{2\pi} d\phi\ \vec{x}\cdot\lambda \\
&= \frac{1}{4}-\frac{1}{8}\vec{x}\cdot\vec{y} = \frac{1}{4}\left(1-\frac{1}{2}\vec{x}\cdot\vec{y}\right)\,,
\end{aligned}
$$

which is what we wanted to prove. □

So we see that the correlations $p_r(00|\vec{x}\vec{y})$ admit a local model for any choice of measurements for $r=\frac{1}{2}$. The similar statement when $r<\frac{1}{2}$ comes from the fact that such a correlation always arises from $p_{\frac{1}{2}}(00|\vec{x}\vec{y})$ by mixing it with uncorrelated random data.

We see then that the Werner states (2.3) with $r\in(0,\frac{1}{2})$ are entangled states that cannot display any nonlocality (via projective measurements). This seminal result by Werner was later improved in several ways. First, Barrett [19] presented a LHV model for POVMs on a Werner state whenever $r<\frac{5}{12}$, providing hence the ultimate proof that entanglement is not a sufficient condition to display nonlocality at single-copy Bell experiments (see next section). Moreover, the model for projective measurements was later improved by Acín *et al.* [20] giving a locally causal explanation for the correlations arising from ρ_r^W with $r<\frac{2}{3}$.

2.2 Quantum nonlocality: activation and hidden nonlocality

So far we have discussed the possibility for a quantum state to generate nonlocal correlations in a traditional (a.k.a. single copy) Bell scenario. But as Popescu noted [21] there are other ways to process a quantum state and generate correlations from it. Some ways to ‘reveal’ the nonlocality from a state involve the following methods:

- Local filtering: locally pre-process each part of the shared state by performing a single measurement of a single outcome, i.e. make each part go through a filter. Then, apply the measurements of the Bell experiment. For a list of examples see Section III-A-2 of [22].

- Multi-copy: The parties share several copies of the same state, and the measurements for the Bell experiment apply to the joint state. Examples of this can be found in [23, 24, 25], among others.
- Networks: joint measurements are made in many copies of the state, distributed in a network configuration. The Multi-copy method is just a particular case of this. An example for multipartite Bell scenarios was first proposed in [26] and will be review later.

In the following I will review the example from [21] that utilises the first method.

Consider a bipartite Bell scenario where Alice and Bob share two qudits. The initial state of the qudits is given by

$$\rho_W^{(d)} = \frac{1}{d^2}\left(\sum_{\substack{i,j\\ i<j}} (|ij\rangle - |ji\rangle)(\langle ij| - \langle ji|) + \frac{1}{d}\mathbb{1}_{d^2}\right) .$$

The state $\rho_W^{(d)}$ is a generalisation of the Werner state for qudits, and the coefficients have been chosen such that the correlations that arise from directly measuring on $\rho_W^{(d)}$ in a Bell experiment have always a local model, for any number of projective[2] measurements and any dimension.

Now we will show that the state can indeed display nonlocality when it's suitable preprocessed before the Bell experiment. The pre-processing comes from both Alice and Bob applying the local filters:

$$F_A = |0\rangle\langle 0| + |1\rangle\langle 1| \ , \quad F_B = |0\rangle\langle 0| + |1\rangle\langle 1| \ ,$$

i.e. they both project their qudit into a qubit subspace. The normalised state after the filtering is given by

$$\tilde{\rho}_W^{(d)} = \frac{(F_A \otimes F_B)\, \rho_W^{(d)}}{||(F_A \otimes F_B)\, \rho_W^{(d)}||} .$$

After the filters are successfully applied[3], Alice and Bob choose and perform one of the two measurements:

$$M_a^{x=0} = \frac{\mathbb{1}_d + (-1)^a\, \tilde{\sigma}_x}{2} , \quad M_a^{x=1} = \frac{\mathbb{1}_d + (-1)^a\, \tilde{\sigma}_z}{2} ,$$
$$M_b^{y=0} = \frac{\mathbb{1}_d + \frac{(-1)^b}{\sqrt{2}}\,(\tilde{\sigma}_x + \tilde{\sigma_z})}{2} , \quad M_b^{y=1} = \frac{\mathbb{1}_d + \frac{(-1)^b}{\sqrt{2}}\,(\tilde{\sigma}_x - \tilde{\sigma}_z)}{2} .$$

[2] There are recent examples of local filtering that reveals the hidden nonlocality of states which admit a LHV model for POVMs. See [27].

[3] The fact that the measurements are chosen after the state has been postselected is crucial not to open the detection loophole. See next section.

where $\tilde{\sigma}_x$ denotes the operator that applies the Pauli matrix σ_x on the two-qubit subspace spanned by $\{|0\rangle, |1\rangle\}$, and similarly for $\tilde{\sigma}_z$.

A straightforward calculation shows that the correlations given by

$$p(ab|xy) = \mathrm{tr}\left\{M_a^x \otimes M_b^y\, \tilde{\rho}_W^{(d)}\right\} \tag{2.4}$$

violate the CHSH inequality whenever $d \geq 5$.

2.3 Experimental quantum nonlocality

2.3.1 Original experiments with photons

Tremendous experimental progress in quantum optics during the 1960s opened the door to possible tests of quantum nonlocality in the laboratory. First, using atomic cascades, it became possible to create pairs of photons entangled in polarization. Second, the polarization of single photons could be measured using polarizers and photomultipliers.

The CHSH inequality can arguably be regarded as the first experimentally testable Bell inequality. Indeed, only three years after their proposal, Freedman and Clauser [28] reported the first conclusive test of quantum nonlocality, demonstrating a violation of the CHSH Bell inequality by 6 standard deviations. Their setup consisted in the following (see Fig. 2.1):

- a polarisation entangled photon pair is produced using a cascade calcium atom decay. Denoting with H and V the horizontal and vertical polarisation, respectively, the state of the two photons is

  $$|\phi_+\rangle = \frac{|HH\rangle + |VV\rangle}{\sqrt{2}},$$

  for the $J = 0 \to 1 \to 0$ decay. The entangled photons have wavelengths 5513Å and 4227Å.
- both arms of the setup (i.e. both parties) are fundamentally equivalent. First, a lens followed by a wavelength filter effectively selects the photons that correspond to the entangled pair. Then, a linear polarizer followed by a single photon detector perform a measurement on the photon polarisation.

The experiment then consisted in measuring:

- $R(\phi)$: coincidence rate for two photon detection, as a function of the angle ϕ between the two polarizers.
- R_1: coincidence rate for two photon detection, when the polarizer in arm 2 is removed. Similarly, R_2.
- R_0: coincidence rate for two photon detection.

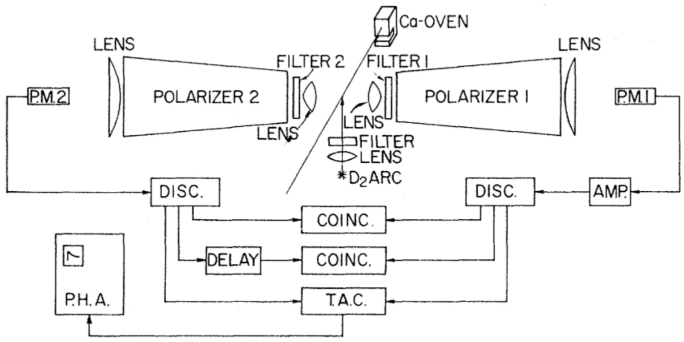


FIG. 1. Schematic diagram of apparatus and associated electronics. Scalers (not shown) monitored the outputs of the discriminators and coincidence circuits during each 100-sec count period. The contents of the scalers and the experimental configuration were recorded on paper tape and analyzed on an IBM 1620-II computer.

Figure 2.1: Freedman and Clauser's Bell-experiment violating local causality [28].

These quantities allow to compute the relative frequencies $\frac{R(\phi)}{R_0}$ and $\frac{R_j}{R_0}$, with which the probabilities are estimated. Now, different relative angles ϕ are related to different pairs of measurement settings by Alice and Bob. Hence, by choosing four different values of ϕ appropriately, one can compute the CHSH value.

Some particulars of this experiment are the following:

- for this particular cascade atom decay, the atom takes away part of the momentum and thus photon directions are not well correlated. This makes the overall detection efficiency to be less than 4%.
- The overall distance covered by each photon since leaving the source until being detected was of the order of meters.
- in practice the setup was static, in the sense that the polarization analyzers were held fixed, so that all four correlations terms had to be estimated one after the other.

Even though these three facts were the state of the art at the moment, they will open the door to loopholes as we will soon see.

A similar experiment done by Aspect *et al.* [29] almost a decade later. The

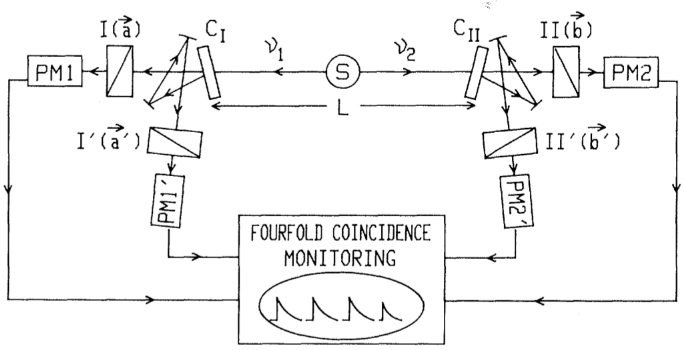


FIG. 2. Timing experiment with optical switches. Each switching device ($C_{\mathrm{I}}, C_{\mathrm{II}}$) is followed by two polarizers in two different orientations. Each combination is equivalent to a polarizer switched fast between two orientations.

Figure 2.2: Aspect, Dalibard and Roger's Bell-experiment violating local causality while closing the locality loophole [30].

relevance of this experiment was not any crucial improvement on itself, but the fact that it allowed for the development of the experiment of [30] a year later. In a nutshell, [29] improves on the production rate of the entangled pair of photons, by selectively pumping the calcium atoms to the upper level of the cascade from the ground state by two-photon absorption. Hence, they attain a better statistical accuracy for the violation.

The main conceptual breakthrough was however achieved by [30], who performed the first Bell experiment with time-varying polarization analyzers. The settings where changed during the flight of the particle and the change of orientation on one side and the detection event on the other side were separated by a space-like interval (see Fig. 2.2), thus closing the locality loophole (see next section). It should be noted though that the choice of measurement settings was based on acousto-optical switches, and thus governed by a quasi-periodic process rather than a truly random one. Nevertheless the two switches on the two sides were driven by different generators at different frequencies and it is very natural to assume that their functioning was uncorrelated [22]. The experimental data turned out to be in excellent agreement with quantum predictions and led to a violation of the CHSH inequality by 5 standard deviations.

This experiment [30] represented the final result of the series of cascade atomic decays ones and allowed to substantially close the space-like loophole. Nevertheless, the collection efficiency was very low, mainly due to the necessity of reducing the divergence of the beams in order to get a good switching [31]. Thus, detection loophole was very far from being eliminated.

2.3.2 Alternative setups

An alternative to these setups based on photons are Bell experiments conducted with atomic systems. Such systems offer an important advantage from the point of view of the detection, with efficiencies typically close to unity. Therefore, atomic systems are well-adapted for performing Bell experiments free of the detection loophole. Rowe *et al.* [32] performed the first experiment of this kind, using two ^{9}Be$^+$ ions in a magnetic trap (see Fig. 2.3). By a coherent stimulated Raman transition, two levels of the ground state are coupled, effectively preparing with fidelity $\sim 80\%$ the state $|\Psi_+\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$ [31]. The "measurement" stage of the Bell experiment consist then in applying a phase to each ion via a a Raman pulse of short duration, whose value corresponds to the "Bell setting", and finally probing each ion with circularly polarised light from a 'detection laser beam'. During this detection pulse, ions in the state $|1\rangle$ scatter many photons, whilst ions in the state $|0\rangle$ scatter very few photons. For two ions one can have three cases: zero ions bright, one ion bright, two ions bright. In the one-ion-bright case Bell's measurement requires only knowledge that the states of two ions are different and not which one is bright. The measured CHSH inequality violation in this setup was $I \sim 2.25 \pm 0.03$, in disagreement with a locally causal model.

The problem with these types of experiments is that, even if they allow for extremely high preparation and detection efficiencies, the measurements of the ions are not performed in a space-like separated manner. Hence, we run into the locality loophole issues we encountered in the first photonic realisations of Bell experiments. For instance, in [32] the ions were located in the same ion trap, separated only by $3\mu m$. This particular issue was considerably improved in 2012 [33] where the ions were separated by $20m$, which is however still far from the minimum required separation of $300m$ for these detection speeds. The novelty of [33] is that there the entanglement between the distant atoms is achieved using an 'event-ready' scheme (see Fig. 2.4): each atom is first entangled with an emitted photon. These two photons are then submitted to a Bell measurement. Upon successful projection of the two photons onto a Bell state, the two atoms become entangled. The scheme is therefore 'event-ready', which makes it robust to photon losses in the channel. Only after the successful Bell measurement, i.e. a successful preparation procedure, the pertinent measurements for the Bell test are conducted.

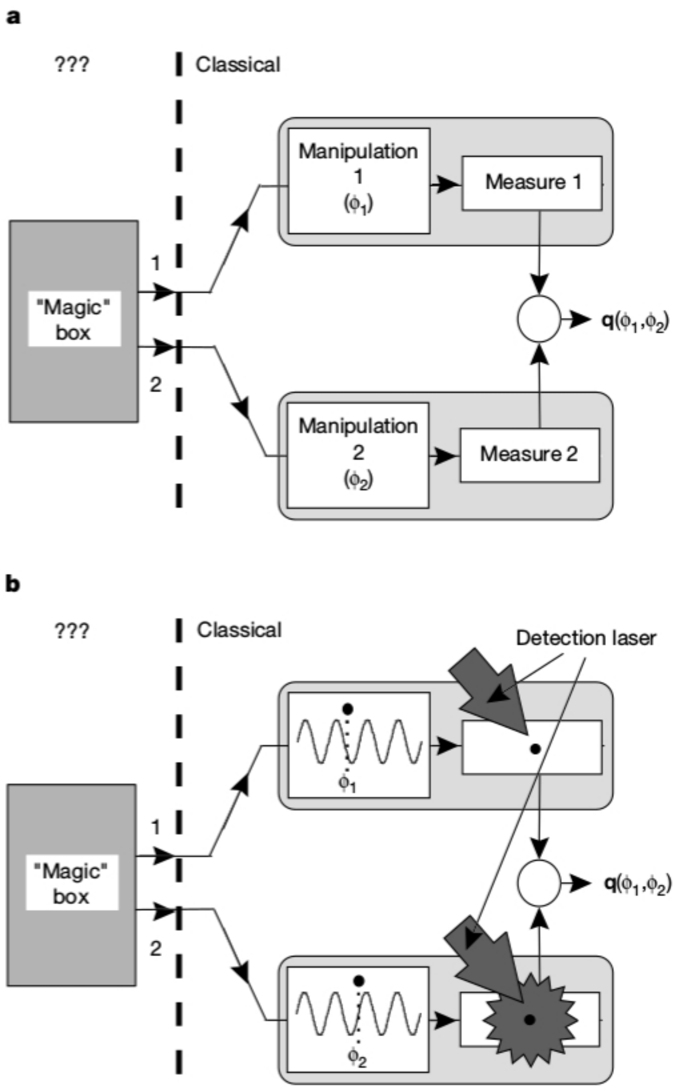


Figure 1 Illustration of how Bell's inequality experiments work. The idea is that a 'magic box' emits a pair of particles. We attempt to determine the joint properties of these particles by applying various classical manipulations to them and observing the correlations of the measurement outputs. **a**, A general CHSH type of Bell's inequality experiment. **b**, Our experiment. The manipulation is a laser wave applied with phases ϕ_1 and ϕ_2 to ion 1 and ion 2 respectively. The measurement is the detection of photons emanating from the ions upon application of a detection laser. Two possible measurement outcomes are possible, detection of few photons (as depicted for ion 1 in the figure) or the detection of many photons (as depicted for ion 2 in the figure).

Figure 2.3: Rowe *et al.*'s Bell-experiment violating local causality while closing the detection loophole [32].

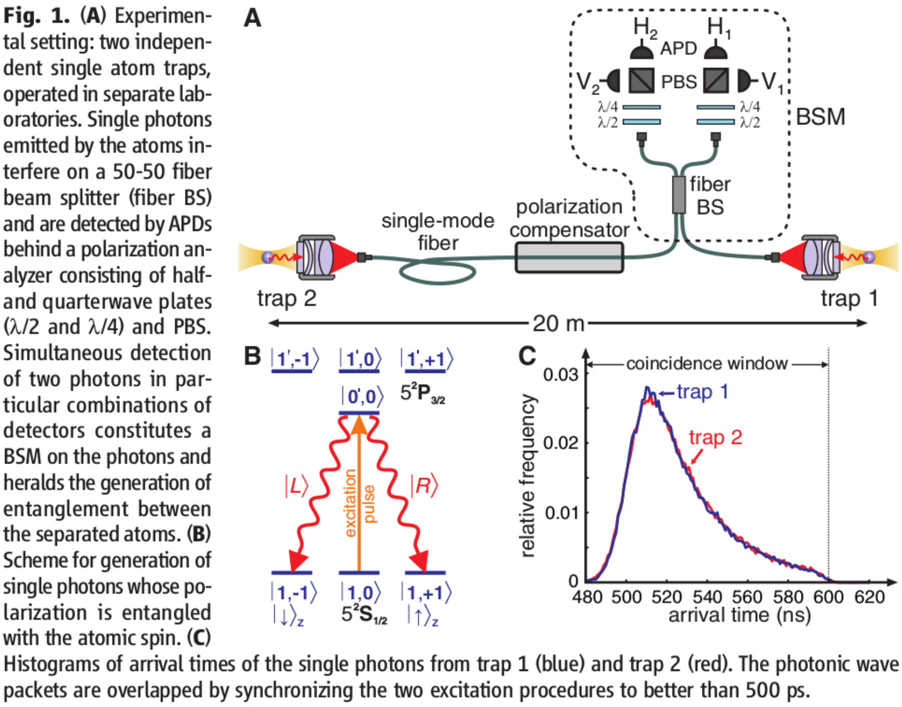


Figure 2.4: Hofmann *et al.*'s Bell-experiment violating local causality while closing the detection loophole using a heralded preparation via an 'event-ready' scheme [33].

2.3.3 Loopholes

The study of loopholes basically asks the question of to which extent alternative models have been falsified by the realised Bell experiments. That is, could the statistical data obtained by them be reproduced by a locally causal model that profits from the experimental imperfections to mimic Bell inequality violations? From a fundamental point of view, there are three main loopholes that need to be addressed in order to answer such question in the negative. These are the following:

- Locality loophole: one of the assumptions in a Bell experiment is that the distant parties perform space-like separated actions. That is, the choice of setting on Alice's side must be space-like separated from the end of the measurement on Bob's side, and vice-versa. This requires that the one should arrange the timing properly, otherwise the detections may be attributed to a LHV model that uses sub-luminal signal. Now, the "end of a measurement" is one of the most fuzzy notions in quantum theory. Consider a photon impinging on a detector: when does quantum coherence leave place to classical results? Already when the photon generates the first photo-electron? Or when an avalanche of photo-electrons is produced? Or when the result is registered in a computer? There is still no consensus, or clear understanding, on when the measurement process ends. Even more, there is an interpretation of quantum theory (Everett's, also called many-worlds) in which no measurement ever happens, the whole evolution of the universe being just a developing of quantum entanglements. All these options are compatible with our current understanding and practice of quantum theory, and give rise to the "quantum measurement problem". As long as this is the situation, strictly speaking it is impossible to close the locality loophole. However, many physicists adopt the reasonable assumption that the measurement is finished "not too long time" after the particle impinges on the detector. With this assumption, the measurement should be finished in the order of microseconds, and hence a distance of 300m between the parties would allow for the loophole to be closed.
- Detection loophole: In all experiments, the violation of Bell's inequalities is measured on the events in which both particles have been detected. However, in a large class of Bell experiments (in particular those carried out with photons), measurements do not always yield conclusive outcomes. This is due either to losses between the source of particles and the detectors or to the fact that the detectors themselves have non-unit efficiency. The simplest way to deal with such 'inconclusive' data, i.e. measurement rounds where some detection process was inconclusive, is simply to discard them

and evaluate the Bell expression on the subset of 'valid' measurement outcomes. The detection loophole hence assumes a form of conspiracy, in which the undetected particle "chose" not to be detected after learning to which measurement it was being submitted. If the detection efficiency is altogether not too high, it is then pretty simple to produce an apparent violation Bell's inequalities with local variables.

As an example, let us see how to fake a violation of the CHSH inequality (1.5) with a LHV model. The model is the following: let the hidden variable λ be the collection of the classical bits x_{guess} and a, i.e. $\lambda = (x_{\text{guess}}, a)$. Now, given the measurement setting y the device on Bob's side will output the bit $b = x_{\text{guess}}\, y \oplus a$. On the other hand, given the measurement setting x, the device on Alice's side will output a whenever $x = x_{\text{guess}}$, and 'no click' otherwise. Note that whenever $x \neq x_{\text{guess}}$, that measurement round will be discarded when post-processing the data to generate the statistics. The correlations that Alice and Bob produce when measuring on a classical system whose underlying description is given by this particular LHV model, are $p^*(ab|xy) = \frac{1}{2}\,\delta_{a\oplus b = xy}$. This correlation gives a CHSH value of $I(p^*) = 4$, which is greater than 2 and hence violates (1.5). In this setup, the detection efficiency in Alice's measurement device is 50%, since when her measurement setting is chosen at random only half of the time it will coincide with x_{guess}.

In order to guarantee that LHV models cannot pull out such strategies to fake Bell inequalities violations, a minimum value for efficiency for the setup is required. The precise minimum value for the efficiencies required to close the detection loophole, depends generally on the number of parties, measurements and outcomes involved in the Bell test. Later on the course we will discuss techniques to compute them and recover some historical thresholds.

- Finite statistics loophole: Since it is expressed in terms of the probabilities for the possible measurement outcomes in an experiment, a Bell inequality is formally a constraint on the expected or average behavior of a local model. In an actual experimental test, however, the Bell expression is only estimated from a finite set of data and one must take into account the possibility of statistical deviations from the average behavior. The conclusion that Bell locality is violated is thus necessarily a statistical one. In most experimental papers reporting Bell violations, the statistical relevance of the observed violation is expressed in terms of the number of standard deviations separating the estimated violation from its local bound. Their are several problems with this analysis, however. First, it lacks a clear operational significance. Second, it implicitly assumes some underlying Gaussian distribution for the measured systems, which is only justified if the number of trials approaches infinity. It also relies on the assumption that the random process associated to the k^{th} trial is independent from the chosen settings and observed outcomes of the previous $k-1$ trials. In other words, the devices are assumed to have no memory, which is a questionable assumption.

A better measure of the strength of the evidence against local models is given by the probability with which the observed data could have been reproduced by a local model. For instance, consider the CHSH test and let $\langle E_{xy}\rangle_{obs}$ be the mean of the observed correlators when measurements x and y are chosen computed over N trials. The probability that two devices behaving according to a local model give rise to a value $I(p)_{obs} = \langle E_{00}\rangle_{obs} + \langle E_{01}\rangle_{obs} + \langle E_{10}\rangle_{obs} - \langle E_{11}\rangle_{obs} \geq 2+\epsilon$ in this finite number of rounds, is given by [34] prob$(I(p)_{obs} \geq 2+\epsilon) \sim e^{-4N\left(\frac{\epsilon}{16}\right)^2}$. This figure of merit is related to the "p-value" used in the experimental papers. This statement assumes that the behavior of the devices at the k^{th} trial does not depend on the inputs and outputs in previous runs. But this memoryless assumption can be avoided and similar statements taking into account arbitrary memory effects can be obtained [22].

Finally, there's another loophole called the "free will loophole" or "superdeterminism loophole", which addresses the possibility that the measurement settings are not chosen at random and that everything is already specified by the unitary evolution of the whole universe. This loophole can never be closed, but even if that's the case whether we should care about it or not is already specified by the evolution unitary ;)

2.3.4 The loophole free ones!

The main limitation in the photonic experiments of the 80s was the low detection rate. Better photodetectors as well as a better source of entangled photon pairs were therefore needed for progressing towards a conclusive experiment. The latter was achieved in the 90s, when spontaneous parametric down conversion in nonlinear crystals became largely exploited. But it wasn't until very recently that the required detectors were finally developed. On the other hand, Bell experiments with ions face the challenge of separation between the magnetic traps, which was also recently surpassed. In this section I will briefly mention the three loophole free Bell tests that were finally realised on 2015.

- The first loophole-free Bell test to appear online was by Hensen *et al.* [35]. In their setup, each party holds a a diamond chip, and the property they measure for the Bell test is the electronic spin m_s associated with a single nitrogen-vacancy in the diamond. The six electrons in the 'vacancy' group together to form an effective spin-1 system. These NV centers naturally occur in diamonds, and their density is very low.

To initialise the spin, light that is resonant to a transition involving $m_s = \pm 1$ states is shone on the system. When the spin comes down to the ground state it may decay to the $m_s = \pm 1$ or to $m_s = 0$. This procedure is repeated until the system is on the $m_s = 0$ state, where it no longer gets excited by the pump. This state will correspond to the "spin up" state for our virtual qubit. Now, to entangle the state of the two NV-centers, light is shone into each diamond resonant to the $m_s = 0$ state. When the state decays back to the ground state, it comes back to $m_s = 0$ and emits a photon. Hence, we have entanglement between the presence of an emitted photon and the electronic spin of the NV center. Now, the photons that come out of Alice and Bob's diamonds meet at a third location, where an even-ready setup performs a measurement on them. Hence, by conditioning on only one photon being detected, the electronic spin of the two nitrogen-vacancies become entangled in a heralded way. The final state for the joint NV vacancies system is a singlet with fidelity $83\% - 96\%$.

For the measurement stage, the experiment goes as follows. First, notice that the transition of this system from the ground state to the excited optical state depends on its the electronic spin (when the temperature is below 8°K). Hence, to read out the spin of the system, they shine a laser that is only resonant with $m_s = 0$. If the NV center is in a bright state ($m_s = 0$) many photons will be emitted and recorded, while when $m_s = \pm 1$ the NV center will remain dark. The choice of measurement setting comes from a quantum random number generator, that chooses one of two different microwave pulses to rotate the spin of the system. This rotation plus the single basis read-out complete the measurement stage, effectively implementing a measurement of the Z basis and the X basis.

They find a value for the CHSH inequality of $I = 2.42 \pm 0.20$, which signals a violation of LC. The p-value of the experiment is 0.019, and goes up to 0.039 allowing classical models with memory. They ran 245 trials of the Bell test during a total measurement time of 220 hours over a period of 18 days.

Note that the detection loophole was avoided by definition, since they are using heralded preparation. The "free-will" loophole was addressed by implementing quantum random number generators, and the locality loophole by taking the parties separated enough to satisfy the appropriate space-time diagrams. The crucial part in the latter was to account for the long duration of the read-out procedure in the measurement stage, which lower bounded the distance between the parties by 1km.

- The second and third loophole-free Bell experiments appeared online simultaneously. Here I will comment on the one by Giustina *et al.* [36]. Their paper reports a violation of a Bell inequality using polarization-entangled photon pairs, high-efficiency detectors, and fast random basis choices spacelike separated from both the photon generation and the remote detection.

  The source distributed two polarization-entangled photons between the two identically constructed and spatially separated measurement stations Alice and Bob (distance ~ 58m), where the polarization was analyzed. It employed type-II spontaneous parametric down-conversion in a periodically poled crystal (ppKTP), pumped with a 405 nm pulsed diode laser. After a photon pair is emitted by the crystal, the photons are coupled into two single-mode optical fibers that direct one photon each to Alice's and Bob's distant locations.

  For the measurement stage, one of two linear polarization directions was selected for measurement, as controlled by an electro-optical modulator (EOM), which acted as a switchable polarization rotator in front of a plate PBS. Customized electronics sampled the output of a random number generator (RNG) to trigger the switching of the EOM.

  The idea to close the detection loophole relies then on (i) the use of high-efficiency detectors, and (ii) the study of a different Bell inequality. This inequality applies to a three-outcome Bell scenario, where one of the outcomes corresponds to the 'no-click' event $\emptyset$, and reads

  $$J = p(++\emptyset|00) - p(+\emptyset|01) - p(\emptyset+|10) - p(++|11) \leq 0\,, \qquad (2.5)$$

  where $+$ denotes a 'click', and may be thought of as the coarse-graining of the two 'real' outcomes of the polarization measurement. To violate this inequality (which automatically closes the detection loophole) efficiencies above 66% are required. In this setup the efficiency of Alice's arm was 78.6% and that of Bob's was 76.2%.

  The source prepared states of the form:

  $$|\Psi\rangle = \frac{|V\rangle\,|H\rangle + r\,|H\rangle\,|V\rangle}{\sqrt{1+r^2}}\,.$$

  The maximum violation of the inequality by such states is $J = 4 \times 10^{-5}$.

In this experiment, they prepare a state with $r \sim -2.9$ and measured at angles $x_0 = 94.4°$, $x_1 = 62.4°$, $y_0 = -6.5°$, and $y_1 = 25.5°$ for approximately 3510 seconds. They obtained a value of $J = 7.27 \times 10^{-6}$, which signals violation of LC since it is a positive quantity. Given that the number is small, it may seem suspicious that it is actually considered a 'significant' violations. But when the p-value for the experimental data is computed, the value found is 3.74×10^{-31}. Hence the probability that a LC model reproduces this finite-statistic experiment is almost null.

- Finally, the third experiment was done by Shalm *et al.* [37]. The setup in this experiment is similar in spirit to that of Giustina *et al.* [36], and also tests the ineq. (2.5). They do have a more intricate way of generating their random inputs, since Alice and Bob each have three different sources of random bits that undergo an `XOR` operation together to produce their random measurement decisions. In one of these sources Alice and Bob each have a different predetermined pseudorandom source that is composed of various popular culture movies and TV shows, as well as the digits of π, and the random bit is generated by processing together all these through an `XOR` gate.

  To test the inequality, they prepare the state

  $$|\Psi\rangle = 0.961\,|HH\rangle + 0.276\,|VV\rangle$$

  and measure along the directions $x_0 = 4.2°$ and $x_1 = -25.9°$ for Alice, and $y_0 = -4.2°$ and $y_1 = 25.9°$ for Bob.

  They report the results from the final data set that recorded data for 30 minutes. There, the best trial gives a violation that has a p-value of 9.2×10^{-6}, hence ruling out LC.

3 Beyond quantum nonlocality: mathematical framework

En esta unidad se presenta la noción de correlationes más allá de lo que la mecánica cuántica puede explicar. Primero se presenta la pregunta de Popescu y Rohrlich, y se introducen las cajas "PR". Luego, se presenta el formalismo para estudiar correlaciones no-signalling, y así el politopo local, el conjunto cuántico, y la interpretación geométrica de las desigualdades de Bell. Al final se mencionan las posibles consecuencias (físicas y en teoría de la información) de la existencia de estas correlaciones más poderosas que las cuánticas.

3.1 Nonlocality beyond quantum

So far we have discussed the notion of a Bell experiment, the constraints that correlations compatible with a locally causal description of reality should satisfy, and how quantum mechanics violates them. In particular, this was first presented by studying the CHSH Bell inequality, whose local bound is given by 2 and which quantum correlations can yield values up to $2\sqrt{2}$. In this case study, one can see that actually the maximum algebraic value of the CHSH expression (1.3) is however 4. Hence a natural question is why quantum correlations cannot realise the values in $(2\sqrt{2}, 4]$. In other words, what constrains quantum correlations to be less nonlocal than what is mathematically allowed.

The first paper to pose this formulation of the problem was by Popescu and Rohrlich [18], who asked whether these constraints could arise from physics. In particular, they wondered if the principle of relativistic causality could render postquatum violations of CHSH unphysical. The precise formulation of this No Signalling principle is as follows:

Definition 3.1. ***No Signalling principle.***
A correlation $p(ab|xy)$ in a Bell scenario satisfies the No Signalling principle if their marginals are well defined, in the sense that the marginal distribution for Alice's

variables is independent of Bob's measurement choice, and vice-versa. Formally,

$$p_A(a|x) \equiv p(a|x,y) = \sum_b p(ab|xy)\,, \quad \forall\, a,x,y\,, \tag{3.1}$$

$$p_B(b|y) \equiv p(b|x,y) = \sum_a p(ab|xy)\,, \quad \forall\, b,x,y\,. \tag{3.2}$$

Correlations that satisfy this principle are called nonsignalling, and define the no-signalling set $\mathcal{NS}$.

What Popescu and Rohrlich noticed is that there exist nonsignalling correlations that can violate the CHSH inequality up to its maximum algebraic value. One example of such, which is usually referred to a PR box[1], is given by the following:

$$p_{\mathrm{PR}}(ab|xy) = \begin{cases} \frac{1}{2} & \text{if} \quad a \oplus b = xy\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{3.3}$$

where $a,b,x,y \in \{0,1\}$ and the sum is taken $\bmod 2$.

It is easy to check that $p_A(a|x) = \frac{1}{2} = p_B(b|y)$, hence the correlations are indeed nonsignalling. Moreover, the correlators have the form $E_{xy} = (-1)^{xy}$, hence $|E_{00} + E_{01} + E_{10} - E_{11}| = 4$.

We see hence that the No Signalling principle alone is not enough to fully characterise the strength of quantum nonlocality. This question since then started a new area of research to try and characterise quantum correlations, their extent and limits, from physical and information theoretical principles. In what follows we will first review a useful mathematical framework to study correlations in Bell experiments, and then move on to discussing the state of the art regarding correlations axioms.

3.2 Probability space

The starting point of the formalism is the identification of each conditional probability distribution $p(ab|xy)$ in an $(2,m,b)$ Bell scenario with a point in probability space.

There are many equivalent ways to define the probability space used to achieve such a representation. One possibility is to consider the vector space $\mathbb{R}^{(md)^2}$, and define the set of allowed correlations as those that consist of positive elements and are well normalised, i.e. those vectors $\vec{p} = [p(11|11),\ldots,p(dd|mm)]$ that satisfy $p(ab|xy) \geq 0 \quad \forall\, a,b,x,y$ and $\sum_{ab} p(ab|xy) = 1 \quad \forall\, x,y$.

[1] Actually, the PR-box was originally discovered by Tsirelson, see eq. (1.11) in [38].

Another possibility is to consider the space with the minimum dimension required so that such a representation is possible. In this case, since for each (x, y) the corresponding distribution is normalised, the value of $p(dd|xy)$ is already fixed by that of the other outcomes. Hence, the probability space can be considered as $\mathbb{R}^{m^2(d^2-1)}$. For instance, in the CHSH scenario the probability vector would live in a 12-dimensional real vector space and its components would read

$$\begin{aligned}\vec{p} = [&p(00|00), p(10|00), p(01|00), p(00|10), p(10|10), p(01|10),\\ &p(00|01), p(10|01), p(01|01), p(00|11), p(10|11), p(01|11)].\end{aligned}$$

The set of allowed correlations is hence here defined by the positivity constraints $\vec{p}(k) \geq 0 \quad \forall\, k \in [1, m^2(d^2-1)]$ and $1 - \sum_{\substack{a,b \\ (ab)\neq(dd)}} p(ab|xy) \geq 0 \quad \forall\, x, y$.

The set of allowed (i.e. well defined) correlations is usually referred to as the *Signalling set*, since there are no constraints imposed on their marginals. Operationally, this set of behaviours will include those in which Alice's and Bob's actions are not space-like separated events.

3.3 The No signalling set

Since the beginning we have stressed that in Bell experiments we usually assume that the parties perform space-like separated actions. So it is only natural to want to focus the study to those correlations that do not allow the parties to signal. That is, starting from the signalling set, we want to further impose the list of linear constraints that come from the No Signalling principle.

First, notice that the No signalling set $\mathcal{NS}$ is hence a polytope, since it arises as the intersection of a finite number of half-spaces[2] (see Fig. 3.1). Second, now that we want to restrict ourselves to the nonsignalling space, the probability vectors can actually be specified by an element in a vector space of smaller dimension. This was first noticed by Collins and Gisin [39], and is usually referred to as Collins-Gisin (CG) form. CG hence expresses the whole conditional probability distribution as a function of the values of $p(ab|xy)$ and its marginals when the outcomes take only the values $\{1, \ldots, d-1\}$. Hence, the minimum dimension of a real vector space where to embed the $\mathcal{NS}$ set is $(1 + m(d-1))^2$. For instance, for a CHSH scenario

[2]Strictly speaking, that condition alone only restricts the set to be either a polytope or a cone. Note, however, that cones are closed under addition and further contain the null vector (i.e. the origin). Since the normalisation constraints on the correlations prevent the $\mathcal{NS}$ set to have either of these properties, $\mathcal{NS}$ is rendered a polytope.

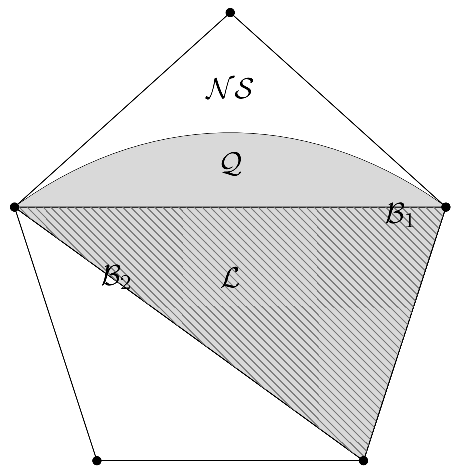


Figure 3.1: Schematic representation of the sets of no-signaling ($\mathcal{NS}$ – pentagon), quantum ($\mathcal{Q}$ – gray area) and classical correlations ($\mathcal{L}$ – striped area). The lines $\mathcal{B}_1$ and $\mathcal{B}_2$ separating the set of classical correlations from the nonlocal ones are examples of tight facet-defining Bell inequalities. While $\mathcal{B}_1$ is violated by some quantum correlations, $\mathcal{B}_2$ is only violated by postquantum nonlocal conditional probability distributions.

a no-signalling probability vector is represented by

$$\vec{p} = [1, p_A(0|0), p_A(0|1), p_B(0|0), p(00|00), p(00|10), p_B(0|1), p(00|01), p(00|11)]\,, \tag{3.4}$$

and the $\mathcal{NS}$ polytope defined by the positivity constraints:

$$\begin{aligned} p(00|xy) &\geq 0 \quad \forall x, y\,, \\ p_A(0|x) - p(00|xy) &\geq 0 \quad \forall x, y\,, \\ p_B(0|y) - p(00|xy) &\geq 0 \quad \forall x, y\,, \\ 1 - p_A(0|x) - p_B(0|y) + p(00|xy) &\geq 0 \quad \forall x, y\,. \end{aligned}$$

The CG notation is particularly useful when implementing optimisation algorithms, both because of its parametrisation and its small dimension.

3.4 The local polytope

A first natural question is that of identifying in probability space the set of correlations that have a locally causal model. This set is usually called *local set*, and here we will denote it by $\mathcal{L}$. As discussed in Section 1.3, these correlations can be equivalently characeterised as those having a local deterministic model. In other words, a correlation is LC if and only if it can be written as a convex combination of locally deterministic conditional probability distributions.

For a fixed Bell scenario $(2, m, d)$, there is a finite number of such deterministic correlations. More precisely, d^{2m}. From a geometrical scope, these define d^{2m} points in probability space, and the set $\mathcal{L}$ is defined by their convex hull. For instance, in the CHSH scenario an example of a deterministic point is that where the parties output 0 regardless of the input. Such a behaviour in CG notation has the form:

$$\vec{p}_D = [1, 1, 1, 1, 1, 1, 1, 1, 1].$$

A deterministic point where Alice always outputs 0 and Bob always 1 instead looks like:

$$\vec{p}_D = [1, 1, 1, 0, 0, 0, 0, 0, 0].$$

Since the number of extreme points is finite for any Bell scenario, $\mathcal{L}$ is a polytope, and hence sometimes is referred to as *Bell polytope*. Fig. 3.1 depicts the local set (among others) in probability space.

Testing whether a correlation is compatible with a locally causal view of reality can be cast as a linear optimisation problem. For small scenarios, one can easily solve this via software optimisations tools, but the complexity of the problem increases exponentially with the size of the Bell experiment, and hence soon becomes computationally intractable. Indeed, it has been shown that this problem is NP-complete[3] [40].

3.5 Facets and Bell inequalities

In the previous section we saw how to characterise the local polytope in terms of its extreme points: i.e. the deterministic behaviours. This is the easiest way to describe the local set, since its vertices are easy to enumerate for any arbitrary scenario. Another equivalent description of the polytope is given by listing its facets, i.e. the hyperplanes whose half-space's intersections defines $\mathcal{L}$. That is, if a point lies below a facets it might admit a LC description, and if not it cannot. Hence, the inequalities defined by the facets are actually Bell inequalities. These Bell inequalities are the ones that were called 'tight' in the strongest sense in

[3] An NP-complete decision problem is one which is both in the NP complexity class and is also NP-hard. The NP class consists of problems whose solutions can be verified efficiently in polynomial time, however there is no known efficient way to find a solution in the first place. A decision problem is NP-hard if, colloquially, it is "at least as hard as the hardest problem in NP".

Section 1.4. For the CHSH scenario, it was proven that all the facets of the local polytope are equivalent to the CHSH inequality [22].

One natural question then is, given a Bell inequality, how to know if it's facet-defining. The answer to this relies on the number of affinely-independent deterministic LC correlations that saturate its bound. Since a facet-defining hyperplane is one dimension smaller than the probability space, then it is saturated by at least a number of affinely-independent deterministic LC correlations equal to the dimension of the probability space. For instance, in the CHSH scenario, the probability space is of effective[4] dimension 8, each CHSH inequality is a hyperplane of dimension 7, and is saturated by 8 deterministic points.

As a final remark, in the most general Bell scenarios not every facet-defining Bell inequality admits a quantum violation, even if they admit one by postquantum nonsignalling ones. Moreover, sometimes the maximum algebraic value of the inequality may already be attained by quantum behaviours even though locally causal ones could not. More on this will be reviewed later on when discussing multipartite Bell scenarios.

3.6 Quantum from principles

The validity of quantum mechanics in the microscopic realm has been established up to incredible precision. However, we still lack physical intuition behind it. This recently motivated a number of works, which have derived the Hilbert space structure of quantum theory from first principles. Their aim is mainly twofold: on the one hand, to legitimize quantum theory, and on the other to explore interesting generalizations of quantum physics by some suitable relaxation of such principles.

This further boosted another line of research where only the correlations that may be observed in Nature are the object we aim to characterise from basic principles. This approach basically takes a step further away by not focusing on the particulars of any physical theory but only on the correlations it may produce. One may naïvely think that this task should be much easier to achieve than that of deriving a full physical theory, but actually is the other way around. As of today we have quite a few different axiomatizations of quantum theory, but we still lack of a purely device independent derivation of their set of correlations. Moreover, it is still unclear whether there are fundamental reasons why such a characterisation may not be possible.

In this section I will discuss the principles that have been proposed so far to constrain the set of accessible correlations in all reasonable physical theories. All of them are satisfies by quantum correlations (and other postquantum ones).

[4]Since the first entry has always value 1 for the normalisation.

No-signalling principle

We have already discussed the No-Signalling principle at the beginning of the lecture, and its formal definition can be found in Def. 3.1.

Non-trivial Communication Complexity

In order to pave the road towards Non-Trivial Communication Complexity (NTCC), let us discuss an example. Consider the case of two distant parties, Alice and Bob, who are each given a string of n bits, $\vec{x}$ and $\vec{y}$ respectively. Their task is to compute the inner product of the two string, $\vec{x}\cdot\vec{y} = \sum_{i=1}^{n} x_i y_i$. For this, they can act locally on some shared correlated system, and moreover exchange some information. NTCC then argues that the amount of communication that is required to successfully perform such a task cannot be trivially small, i.e. when they assist the protocol via shared correlations, these cannot render the amount of required communication trivial.

So consider the case where the parties share n independent copies of a PR box, and perform the following protocol: both Alice and Bob input the ith bit of their respective strings into their respective sides of the ith box, obtaining the outcomes a_i and b_i. Now Alice computes the sum mod 2 of all her outcomes $a = \oplus_{k=1}^{n} a_k$ and sends that bit to Bob. Finally, Bob adds all his outcomes plus the bit a sent by Alice, computing hence

$$\oplus_{k=1}^{n} b_k + a = \oplus_{k=1}^{n} b_k + \oplus_{k=1}^{n} a_k = \oplus_{k=1}^{n}(a_k + b_k) = \oplus_{k=1}^{n} x_k y_k = \vec{x}\cdot\vec{y},$$

succeeding in the task with probability 1.

NTCC hence argues that such a minimal amount of communication allowed but the uses of PR boxes is not ‘sensible’, which renders PR boxes as devices that shouldn’t exist in Nature.

In a more general setup, roughly speaking, the axiom of Non-Trivial Communication Complexity (NTCC) states that two spatially separated parties, call them Alice and Bob, cannot compute arbitrary boolean functions with fixed probability greater than $\frac{1}{2}$ for all input sizes [41]. More concretely, suppose that Alice and Bob are respectively distributed the strings of bits $\overline{x}, \overline{y} \in \{0,1\}^n$. Bob’s task is to compute the function $f(\overline{x}, \overline{y}) \in \{0,1\}$, and, to that effect, Alice is allowed to transmit him one bit of information. For a particular protocol, call $p(\overline{x}, \overline{y})$ the probability that Bob succeeds when the inputs are $\overline{x}, \overline{y}$. NTCC then implies that there exists a family of functions $\{f_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}\}$ such that no communication protocol can succeed with probability $p(\overline{x}, \overline{y}) > p > \frac{1}{2}$ independent of n, for all $\overline{x}, \overline{y} \in \{0,1\}^n$ and all input sizes n. In [41] it is shown that boxes with a Clauser-Horne-Shimony-Holt (CHSH) parameter [6] greater than $4\sqrt{\frac{2}{3}}$ could be used to devise protocols which violate this principle. NTCC thus imposes non-trivial constraints on the set of non-local correlations.

Information Causality

Consider a bipartite scenario similar to the previous one, where Alice receives an n-bit string x and Bob a random number $k \in \{1, \ldots, n\}$. Bob's task consists in making a guess b for Alice's bit x_k. To aid him, Alice is allowed to send Bob $m < n$ bits of information.

Note that, if this protocol could be played perfectly, as soon as Alice sent her m bits, Bob would be in possession of a box that would potentially contain n of Alice's bits. This is because Alice cannot use the information about k when crafting the message m. However, one would intuitively expect Bob's system to hold no more than m potential bits of information. The principle of Information Causality [42] tries to capture this intuition by stating that

$$\sum_{k=1}^{n} I(b : x_k|k) \leq m.$$

Here, $I(A : B)$ denotes the mutual information between the random variables A and B, i.e., $I(A : B) = H(A) + H(B) - H(A, B)$, with the Shannon entropy being $H(Z) = -\sum_Z p_Z \log_2(p_Z)$.

The exact constraints that IC places on the strength of nonlocal correlations are, up to this day, unknown. However, this topic has received considerable attention, and several limitations in different nonlocality scenarios have been established. More importantly, it's been shown that quantum correlations do satisfy the principle.

No Advantage for Nonlocal Computation

Nonlocal computation is an information processing task introduced in [43].

In this primitive, an n-bit string $z \in \{0, 1\}^n$ is distributed with prior probability $\tilde{p}(z) \geq 0$, with $\sum_z \tilde{p}(z) = 1$. The goal behind nonlocal computation is to have two non-communicating parties, Alice and Bob, evaluate the Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ on z while learning nothing about the value of z. For that purpose, a fully random n-bit string x is generated and sent to Alice, while Bob receives the bit string $y \equiv z \oplus x$. Given inputs x, y, Alice and Bob's task is to produce two binary outputs a, b such that $a \oplus b = f(x \oplus y) = f(z)$.

The figure of merit of nonlocal computation is Alice and Bob's average success at computing f, given by the expression

$$P(f) = \frac{1}{2^n} \sum_{xy} \tilde{p}(x \oplus y) p(a \oplus b = f(x \oplus y)|xy) \,.$$

Whenever Alice and Bob can assist their computation protocols by using shared classical correlations, i.e. shared randomness, then the maximum probability of success is [43]:

$$P_C(f) = \frac{1}{2} \left(1 + \max_{u \in \{0,1\}^n} \left| \sum_z (-1)^{f(z)+u \cdot z} \tilde{p}(z) \right| \right) \,.$$

The No Advantage for Nonlocal Computation (NANLC) principle then states that $P(f) = P_C(f)$. I.e. it doesn't matter which correlations available in Nature Alice and Bob share, they may not succeed on this task task better that if they shared classical resources. Quantum correlations do satisfy this principle [43].

Macroscopic Locality

In a nutshell, Macroscopic Locality (ML) states that coarse-grained extensive measurements of N independent particle pairs must admit a local hidden variable model in the limit $N \to \infty$ [44]. ML is justified on the grounds that any reasonable physical theory must have a classical limit; ergo, 'natural' macroscopic experiments should be describable by a classical theory, and consequently their associated statistics must be local realistic.

In this setup, a source S prepares a pair of particles s_A and s_B, which are sent to the measurement devices M_x^A and M_y^B at Alice's and Bob's labs respectively. There, an interaction between the measurement apparatus and the system sends each particle towards one of a set of detectors, where its presence can be observed as a "detector click". The clicking of detector D_k corresponds to obtaining outcome k when measurement M is performed (see Fig. 3.2). This is actually a rephrasing of a Bell experiment, and hence the correlations $p(ab|xy)$ are those which we have been discussing until now.

A macroscopic version of this experiment then consists of a source S that now prepares N independent pairs of particles. Hence, in each round N particles enter each measurement device. The assumption then is that we are no longer able to distinguish individual outcomes, but rather the fraction of instances (or "intensity") of each outcome k given a measurement. Moreover, we also assume that the measurement device can act on each particle independently, just as if the particle had entered it alone (think, for instance, of a Stern-Gerlach type of device). Hence, for each particle (independently), an interaction between the measurement

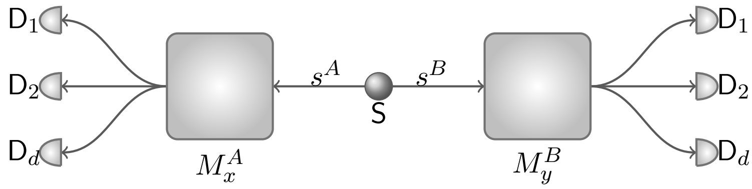


Figure 3.2: Microscopic experiment in the "Macroscopic locality" setup.

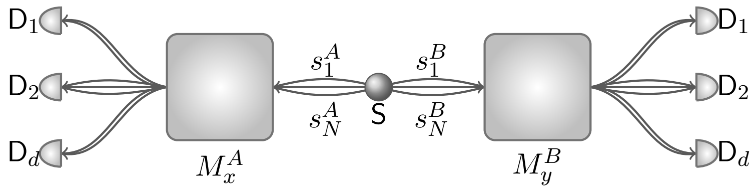


Figure 3.3: Macroscopic experiment in the "Macroscopic locality" setup.

apparatus and it sends the particle towards one of a set of detectors. However, in this case, rather than a single click there is a distribution of 'clicks' over the detectors according to the probabilities for each outcome in the microscopic experiment. Hence, the 'output' of this macroscopic experiment is the collection of intensities I_k^M registered at the detectors (see Fig. 3.3). In this setup then, the correlations that are measured take the form $P(I_a I_b|xy)$.

Macroscopic locality then demands that when $N \to \infty$, $P(I_a I_b|xy)$ has a locally causal model. This hence imposes constraints on how strong the microscopic correlations $p(ab|xy)$ may be.

The ML principle was later on refined by the Macroscopic Noncontextuality (MNC) principle [45], and in a future lecture I will comment on the details and their consequences.

Local Orthogonality

The Local Orthogonality principle (LO) basically imposes constraints on the probabilistic models by means of the orthogonal events of the scenario [46, 47]. LO has applicability as a hierarchy of constraints, which I will further present in the following.

The first ingredient in the LO principle is the notion of *orthogonal events*. An event refers to a measurement round where measurements were performed and outcomes obtained. I.e., each event is labelled by $(ab|xy)$. Basically, LO tells that in a Bell scenario, whenever we consider two events where there exists a party that has performed the same measurement and obtained different outcomes, those two events are rendered orthogonal. For instance, $(00|00) \perp (10|01)$, whereas the orthogonality between $(00|00)$ and $(01|01)$ cannot be determined. In other words, two events are orthogonal if the local assessment of a single party is enough to tell that the two measurement rounds correspond to incompatible situations.

The last ingredient of LO is that

$$\sum_{(ab|xy)\in S} p(ab|xy) \leq 1\,, \tag{3.5}$$

where S is a set of pairwise orthogonal events.

Example 3.2. In a CHSH scenario, the set $S = \{(00|00), (01|00), (10|01), (11|01)\}$ is one of mutually orthogonal events. Hence, one LO constraint in the CHSH scenario reads

$$p(00|00) + p(01|00) + p(10|01) + p(11|01) \leq 1\,.$$

LO then imposes constraints on the correlations p via eq. (3.5) both

- at the single-copy level, applying eq. (3.5) straightforwardly (as in the previous example).
- when k copies of the device are considered and now the constraint

$$\sum_{(a_1b_1\ldots a_kb_k|x_1y_1\ldots x_kb_k)\in S} \otimes_{i=1}^{k} p(a_ib_i|x_iy_i) \leq 1\,, \tag{3.6}$$

is imposed. For each k, we say that the LO principle is applied at level k.

Actually, how to impose LO at higher levels requires multipartite Bell scenarios, which I will discuss next lecture. So hopefully this last point will become clearer then, when I discuss an example.

4 Almost-quantum correlations and multipartite Bell scenarios

En esta unidad se presenta el problema de caracterizar la forma del conjunto de correlaciones cuánticas, en particular su borde. Se presenta primero la jerarquía de relajaciones propuesta por Navascués, Pironio y Acín (NPA) , y luego los principios físicos y de teoría de la información que intentan recuperar el borde. Se introduce entonces el conjunto de correlaciones casi-cuánticas, y se dicute por un lado su viabilidad para describir la naturaleza, y por el otro las posibles limitaciones de este enfoque de investigación. Finalmente se generalizan los escenarios de Bell a situaciones multipartitas. Se presentan las particularidades de estos casos, como ser la monogamía de las correlaciones y las diferentes clases de nolocalidad, haciendo hincapié en las multipartitas-genuinas y desigualdades de Svetlichny.

4.1 Quantum correlations from a hierarchy of semidefinite relaxations

An interesting open problem is that of how to characterise quantum correlations *from the outside*. That is, start from the nonsignalling set and via physical or information-theoretical principles render all postquantum correlations unphysical. Several principles have been proposed, but none of them has succeeded on reproducing the quantum boundary completely.

When the search for such a physically motivated principle is dropped, however, an algorithmic characterisation of the boundary of quantum correlations under the commutativity paradigm was found by Navascués, Pironio and Acín (NPA) [48, 49, 50]. NPA hence provides a method to algorithmically determine membership to $\mathcal{Q}$.

More precisely, NPA provided an algorithmic characterisation of $\mathcal{Q}$ via a hierarchy $\mathcal{Q}_1 \supseteq \mathcal{Q}_2 \supseteq \ldots \supset \mathcal{Q}_k$ of (postquantum) subsets to $\mathcal{NS}$ which converges to $\mathcal{Q}$ in the asymptotic limit of $k \to \infty$ for any Bell scenario. Each set $\mathcal{Q}_k$ (for positive integer k) is characterised by the set of feasible solutions to a semidefinite program (SDP), which demands that a Hermitian matrix Γ_k —whose entries being *all* the moments (expectation values) up to order $2k$—can have only non-negative eigenvalues. Importantly, some of these moments correspond precisely to those joint conditional probabilities between measurement outcomes, that can be estimated directly from a Bell-type experiment. In general, however, the non-negativity of a moment matrix Γ_k is not sufficient to guarantee that the behaviour p has a quantum origin. In other words, for *any* given k, all given p where the corresponding Γ_k (which contains p as entries) can be made non-negative form a superset of $\mathcal{Q}$, which is precisely the NPA set $\mathcal{Q}_k$.

So let me stop making general statements, and let us discuss how these moment matrices Γ_k are defined in bipartite Bell scenarios. Let m be the number of measurements per party, each of which has d possible outcomes. The starting point for each level in the hierarchy is to define the set of *labels* for the rows and columns of the moment matrix Γ_k, and composition rules on how those lables are combined. Each of these labels is composed by an ordered sequence of single-party events, in the following way. For the first level, the set is given by:

$$\begin{aligned} S_1 = \emptyset \cup \{(a|x) \,:\, 1 \leq x \leq m \,,\, 1 \leq a \leq d-1\} \cup \\ \{(b|y) \,:\, 1 \leq y \leq m \,,\, 1 \leq b \leq d-1\} \,. \end{aligned}$$

That is, the labels represent either a 'null' event or single-party events.

Now, for the second level of the hierarchy, the labels can further comprise sequences of two single-party events, i.e.

$$\begin{aligned} S_2 = S_1 \cup \{(aa'|xx') \,:\, 1 \leq x, x' \leq m \,,\, 1 \leq a, a' \leq d-1\} \cup \\ \{(bb'|yy') \,:\, 1 \leq y, y' \leq m \,,\, 1 \leq b, b' \leq d-1\} \cup \\ \{(ab|xy) \,:\, 1 \leq x, y \leq m \,,\, 1 \leq a, b \leq d-1\} \,. \end{aligned}$$

As you might have already realised, the set of labels S_k for level k will comprise sequences of at most k single party events.

Now, these labels relate to each other according to composition/concatenation rules. First, since the measurements for a single party may not commute, $(aa'|xx')$ and $(a'a|x'x)$ do represent different labels, while $(ab|xy)$ and $(ba|yx)$ are rendered equivalent. Moreover, a sequence (or a fraction of one) composed by identical events will be identified with that event alone. I.e. $(abb|xyy) \equiv (ab|xy)$. These rules may seem arbitrary, but actually they arise by thinking how projectors labelled by the words in these sets would behave.

Now that we know how to label the rows and columns of the moment matrix Γ_k

for each level k, we can phrase the constraints that its elements should satisfy:

$$\Gamma_{\emptyset,\emptyset} = 1\,, \tag{4.1a}$$
$$\Gamma_{\emptyset,(a|x)} = p_A(a|x)\,, \tag{4.1b}$$
$$\Gamma_{\emptyset,(b|y)} = p_B(b|y)\,, \tag{4.1c}$$
$$\Gamma_{\emptyset,(ab|xy)} = p(ab|xy)\,, \tag{4.1d}$$
$$\Gamma_{u,v} = \Gamma_{r,s} \quad \text{if} \quad u^\dagger v \equiv r^\dagger s\,, \tag{4.1e}$$

where $u^\dagger$ denotes the sequence u in reverse order.

Now, if for the behaviour p there exists a positive semidefinite matrix Γ_k with rows and columns labelled by S_k and that satisfies conditions (4.1a) to (4.1e), then $p \in \mathcal{Q}_k$. The set $\mathcal{Q}_k$ is hence defined (only) by all the correlations p such that this happens.

Moreover, notice that since the moment matrix Γ_{k+1} associated with the characterisation of $\mathcal{Q}_{k+1}$ contains the moment matrix Γ_k associated with the characterisation of $\mathcal{Q}_k$ as a submatrix, it thus follows that $\mathcal{Q}_{k+1} \subseteq \mathcal{Q}_k$ for all positive integer k.

As I loosely mentioned before, the NPA hierarchy converges when $k \to \infty$ to the set of quantum correlations under the commutativity paradigm.

So far, the NPA hierarchy has served to main purposes. On the one hand, it's a useful test to certify postquantumness of a behaviour p. This is because this task can be solved by the answer to a single test, i.e. failure of membership to a particular level. On the other hand, NPA is useful for providing upper bounds to the Tsirelson's bound of Bell inequalities. Computing the maximum violation of a particular inequality by correlations in $\mathcal{Q}_k$, for a fixed k, is an SDP, which is considered to be an efficiently solvable problem. In some cases, these upper bounds actually converge to the actual Tsirelson's bound in a finite number of steps. This is the case, for instance, of the CHSH inequality. A counter example to this is the I_{3322} inequality, where so far Tsirelson's bound is unknown, and only estimates from a finite number of NPA sets are known [51].

As a comment, for bipartite scenarios, the correlations in $\mathcal{Q}_1$ provably satisfy the Macroscopic Locality principle [44], which is itself a proof that quantum correlations do so as well.

4.2 Almost quantum correlations

Almost quantum correlations form a particular set of nonsignalling correlations that strictly contains quantum behaviours. They have caught the attention of the community since (i) they are easy to characterise via an SDP, like the NPA sets, (ii) they provably satisfy all the principles that have been proposed so far to characterise quantum correlations, with the possible exception being IC for which there's only numerical evidence to support the claim (and not an analytical proof).

There are many equivalent ways to define almost quantum correlations [52], each of which has proven useful in different situations. Here I will comment on the two main ones.

Definition 4.1. ***Almost quantum correlations****.*
A conditional probability distribution $p(ab|xy)$ is almost-quantum if there exists a Hilbert space $\mathcal{H}$, a projective measurement $\{\Pi_{a|x}\}_a$ in $\mathcal{H}$ for each x for Alice, a projective measurement $\{\Pi_{b|y}\}_b$ in $\mathcal{H}$ for each y for Bob, and a quantum state $\rho = |\psi\rangle\langle\psi|$ in $\mathcal{H}$ such that the statistics are recovered by them, i.e.

$$p(ab|xy) = \mathrm{tr}\left\{\Pi_{a|x}\,\Pi_{b|y}\,\rho\right\}\ ,$$

and the measurements operators for different parties commute on the state, i.e.

$$\Pi_{a|x}\,\Pi_{b|y}\ |\psi\rangle = \Pi_{b|y}\,\Pi_{a|x}\ |\psi\rangle\ .$$

The set of all correlation that have an almost quantum realisation defines the almost quantum set, denoted by $\widetilde{\mathcal{Q}}$. Note that if the last condition (commutativity) is demanded on all states in $\mathcal{H}$ and not just that particular ρ, the set that we recover is that of quantum correlations under the commutativity paradigm.

Alternatively, the same set of almost quantum correlations can be characterised as a semidefinite program, in the spirit of the NPA sets previously discussed. Indeed, originally almost quantum correlations were referred to as the '1+AB' level in the NPA hierarchy for bipartite scenarios. The reason for such a name will become clear in what follows.

Definition 4.2. ***Almost quantum correlations as an SDP****.*
A conditional probability distribution $p(ab|xy)$ is almost-quantum if and only if there exists a positive semidefinite moment matrix $\Gamma_{\widetilde{\mathcal{Q}}}$, with rows and columns labelled by

$$S_{\widetilde{\mathcal{Q}}} = S_1 \cup \{(ab|xy)\ :\ 1 \le x, y \le m\ ,\ 1 \le a, b \le d-1\}\ ,$$

that satisfies the linear constraints (4.1a) *to* (4.1e)*.*

We see that for bipartite scenarios, $S_1 \supset S_{\widetilde{\mathcal{Q}}} \supset S_2$, hence the set of almost quantum correlations lies between the first and second levels of NPA. This is no longer true for multipartite Bell scenarios, as we will discuss later on.

4.3 Multipartite Bell scenarios: types of nonlocality

The notion of nonlocality that we have discussed during the past lectures for the case of two distant parties readily extends to three or more parties. In the multipartite case, however, nonlocality displays a much richer and more complex structure compared to bipartite scenarios. This makes the study and the characterization of multipartite nonlocal correlations an interesting but challenging problem. Hence our understanding of nonlocality in the multipartite setting is much less advanced than in its bipartite counterpart.

In general, a multipartite Bell scenario consists on n distant parties that perform space-like separated actions on their share of a physical system. Each party can choose among m_k measurements to perform on their system, each yielding one out of d_k possible outcomes, where $k \leq n$ labels the party. Even though in principle m_k and d_k may differ from party to party, it is possible to embed such situations in scenarios where these numbers all coincide, i.e. $m_k = m$ and $d_k = d$ for all k. Hence, from now on we will identify a multipartite Bell scenario by the numbers (n, m, d).

The study of multipartite nonlocality was initiated by the ground-breaking work of Svetlichny [53]. In this paper, the author introduced the concept of genuine multipartite nonlocality, derived a Bell-type inequality for testing it, and showed that that this strong form of nonlocality occurs in quantum mechanics. Later, in particular with the advent of quantum information science, the concepts and tools introduced by Svetlichny were further developed.

So let us begin by discussing the definition of "genuine multipartite nonlocality". For simplicity we will consider the case of three parties: Alice, Bob and Charlie. Similarly to the bipartite case, each party is assumed to perform space-like separated actions on their share of a system. Their measurement choices and outcomes are labelled by x, y, z and a, b, c for each party respectively. Correlations that have a LC explanation within this setup are hence those of the form

$$p(abc|xyz) = \int d\lambda\, q(\lambda)\, p_A(a|x, \lambda)\, p_B(b|y, \lambda)\, p_C(c|z, \lambda)\,, \tag{4.2}$$

where $q(\lambda)$ is a normalised probability distribution on the classical common cause λ. Similarly to section 1.3, we can assume, without loss of generality, λ to belong to a finite set of "deterministic strategies", and the local distributions p_A, p_B, p_C to be deterministic functions.

In the multipartite case, furthermore, there exist several possible refinements of this notion of nonlocality. For instance, one could have tripartite correlations of the form $p(abc|xyz) = p(ab|xy)p(c|z)$, i.e. Charlie is uncorrelated to Alice and Bob. Such correlations can only exhibit bipartite nonlocality, and that only depends on $p(ab|xy)$. On the other hand, correlations that display a stronger form on nonlocality, i.e. one that relies on the three parties altogether rather than on their bipartitions, are called "genuine" tripartite. Now, the explicit definition of this set is still somehow under debate, as we will see below.

The first notion of genuine multipartite nonlocality is the "à la Svetlichny" [53]. To describe it let us consider the situation where only two parties share a nonlocal resource or communicate in any measurement run. That is, correlations of the form

$$\begin{aligned} p(abc|xyz) = & \int d\lambda\, q(\lambda)\, p_{AB}(ab|x,y,\lambda)\, p_C(c|z,\lambda) \\ & + \int d\mu\, q(\mu)\, p_{AC}(ac|x,z,\mu)\, p_B(b|y,\mu) \\ & + \int d\nu\, q(\nu)\, p_{BC}(bc|y,z,\nu)\, p_A(a|x,\nu)\,, \end{aligned} \tag{4.3}$$

where $\int d\lambda\, q(\lambda) + \int d\mu\, q(\mu) + \int d\nu\, q(\nu) = 1$. This represents a convex combination of three terms, where in each term at most two of the parties are nonlocally correlated. Correlations of this form are called "2-way nonlocal", and form the set $S_{2|1}^{\mathrm{Sve}}$.

Operationally, we can define local correlations as those that can be generated by separated classical observers that have access to share randomness but who cannot communicate, 2-way correlations as those where arbitrary communication is allowed between two parties, and 3-way (or genuine tripartite) correlations as those where arbitrary communication is allowed between all parties.

Svetlichny's definition of genuine multipartite nonlocality, however, is not fully accepted by the community, mainly because there are no constraints on the bipartite correlations p_{AB}, p_{AC}, p_{BC} in eq. (4.3). That is, the potential bipartite nonlocal correlations in eq. (4.3) are allowed to be signalling as long as their effective combination yields a tripartite nonsignalling one. In order to refine this, people suggested to define 2-way correlations as in eq. (4.3) but further demanding that p_{AB}, p_{AC}, p_{BC} be nonsignalling [54, 55]. This set of correlations is usually denoted as $S_{2|1}^{\mathrm{NS}}$. Another refinement was later proposed by Gallego *et al.* [56], who introduced the notion of "time ordered bilocal models". There, the bipartite correlations p_{AB}, p_{AC}, p_{BC} in eq. (4.3) are not asked to be nonsignalling, but rather to display at most one-way signalling. That is, for instance, p_{AB} can either display signalling from Alice to Bob or from Bob to Alice (depending on λ) but not both. I will refer to the set of time-ordered bilocal models, following [22], as $S_{2|1}^{\mathrm{TO}}$. It follows that

$$\mathcal{L} \subset S_{2|1}^{\mathrm{NS}} \subset S_{2|1}^{\mathrm{TO}} \subset S_{2|1}^{\mathrm{Sve}}.$$

It follows that, while violation of Svetlichny's decomposition (4.3) signals genuine tripartite nonlocality, there exist some correlations whose tripartite character only manifests itself when considering the weaker definitions of $S_{2|1}^{\mathrm{NS}}$ or $S_{2|1}^{\mathrm{TO}}$.

Now that we have discussed the challenges on how to properly define genuine multipartite nonlocality and all its intermediate types, let us move on to how to detect it. The first to devise such a test was Svetlichny [53], who derived an inequality that holds for correlations of the type given by eq. (4.3) in a $(3, 2, 2)$ Bell scenario. This inequality reads

$$\begin{aligned} S_3 = & \langle A_1 B_1 C_2 \rangle + \langle A_1 B_2 C_1 \rangle + \langle A_2 B_1 C_1 \rangle - \langle A_2 B_2 C_2 \rangle \\ & + \langle A_2 B_2 C_1 \rangle + \langle A_2 B_1 C_2 \rangle + \langle A_1 B_2 C_2 \rangle - \langle A_1 B_1 C_1 \rangle \underset{\text{2-way}}{\leq} 4 \,. \end{aligned} \tag{4.4}$$

To see that correlations of the form given by eq. (4.3) satisfy this bound, consider first the case where Charlie is uncorrelated, i.e. $p_{AB}(ab|x, y, \lambda)\, p_C(c|z, \lambda)$. Now, rewrite S_3 as

$$\begin{aligned} & (\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle) \, \langle C_2 \rangle \\ & + (\langle A_2 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_1 B_2 \rangle - \langle A_1 B_1 \rangle) \, \langle C_1 \rangle \,. \end{aligned} \tag{4.5}$$

This expression can be understood as follows: depending on Charlie's choice of measurement, Alice and Bob play the CHSH game or its symmetric version (i.e. where the role of the measurement settings in the inequality are interchanged). Now, whenever $(\langle A_1 B_1 \rangle + \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle - \langle A_2 B_2 \rangle) = 4$, i.e. when Alice and Bob share a PR box, $(\langle A_2 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_1 B_2 \rangle - \langle A_1 B_1 \rangle) = 0$, and vice-versa, yielding a value of 4 for eq. (4.5). The same can be proven to hold when Alice and Bob do not share PR boxes. Now, when the uncorrelated party is Alice instead of Charlie, i.e. $p_{BC}(bc|y, z, \nu)\, p_A(a|x, \nu)$, Bob knows which CHSH game he's supposed to play with Alice, since he is though of as being together with Charlie. However, since here Alice and Bob are uncorrelated they cannot do better at the CHSH game than yielding a value of 2, hence eq. (4.5) achieves a maximum value of 4. The same can be concluded when Bob is the uncorrelated party in (4.3), which proves the bound in eq. (4.4).

Quantum mechanics can, however, violate inequality (4.4). For this, consider the case where the three parties share three qubits prepared on the GHZ quantum state $|\mathrm{GHZ}\rangle = \frac{|000\rangle + |111\rangle}{\sqrt{2}}$. Let Alice's, Bob's and Charlie's measurement operators be given by:

$$M_a^{x=0} = \frac{\mathbb{1} + (-1)^a \, \sigma_x}{2} \,, \quad M_a^{x=1} = \frac{\mathbb{1} + (-1)^a \, \sigma_y}{2} \,,$$
$$M_b^{y=0} = \frac{\mathbb{1} + \frac{(-1)^b}{\sqrt{2}} \, (\sigma_x - \sigma_y)}{2} \,, \quad M_b^{y=1} = \frac{\mathbb{1} + \frac{(-1)^b}{\sqrt{2}} \, (\sigma_x + \sigma_y)}{2} \,,$$
$$M_c^{z=0} = \frac{\mathbb{1} - (-1)^c \, \sigma_y}{2} \,, \quad M_c^{z=1} = \frac{\mathbb{1} + (-1)^c \, \sigma_x}{2} \,,$$

where σ_x, σ_y and σ_z are, as usual, the Pauli matrices. That is, Alice measures the observables $A_1 = \sigma_x$ and $A_2 = \sigma_y$, Bob the observables $B_1 = \frac{\sigma_x - \sigma_y}{\sqrt{2}}$ and $B_2 = \frac{\sigma_x + \sigma_y}{\sqrt{2}}$, and Charlie $C_1 = -\sigma_y$ and $C_2 = \sigma_x$. These correlations achieve a value for eq. (4.4) of $S_3 = 4\sqrt{2} > 4$.

For generalisations of Svetlichny's inequality to more parties, measurements and dimensions and further results on genuine multipartite nonlocality see [22] and references therein.

4.4 Monogamy

The monogamy of nonlocal correlations is a property that appears in multipartite Bell scenarios. To illustrate it, let us consider the CHSH inequality in a $(3, 2, 2)$ scenario. More precisely, let us consider both the CHSH expression (1.3) between Alice and Bob, here referred to as I_{AB}, and the one between Alice and Charlie, here referred to as I_{AC}. The question is whether Alice can violate the CHSH inequality with both Bob and Charlie at the same time when sharing a nonsignalling correlation $p(abc|xyz)$ with them.

When the parties are allowed to share any arbitrary quantum correlations, i.e. for the general case of an arbitrary quantum state shared by the three parties, this question was answered in the negative by [57], who showed that $I_{AB}^2 + I_{AC}^2 \leq 8$. This implies, for instance, that if Alice and Bob violate the CHSH inequality maximally, then Alice and Charlie must be uncorrelated.

This "monogamy" of correlations, however, is not specific to the CHSH inequality but applies to essentially all bipartite Bell inequalities. Moreover, even nonsignalling correlations are surprisingly monogamous [58]. In what follows, I will present the monogamy results of [59] as an example.

Consider then an arbitrary Bell inequality in a $(2, m, d)$ scenario given by

$$I(p) = \sum_{a,b,x,y} \alpha(a, b, x, y) \, p(ab|xy) \leq \beta \,, \tag{4.6}$$

where without loss of generality we can take $\alpha(a,b,x,y) \geq 0$ for all a,b,x,y. Hence, $0 \leq I(p) \leq \beta$.

The goal is to study the monogamy relations given by such an inequality in a (n,m,d) Bell scenario with $n = m+1$. For the purpose of the narrative, we consider these n parties to be one Alice and m Bobs. Now denote by I_k the value of the Bell expression (4.6) given by the correlations between Alice and the k$^{\text{th}}$ Bob. The monogamy relation that we will show nonsignalling correlations to satisfy is given by

$$\sum_{k=1}^{m} I_k \leq m\beta. \tag{4.7}$$

To prove the statement we will see that a violation of (4.7) would imply signalling. For this, notice that the LHS of eq. (4.7) can be rewritten as

$$\sum_{k=1}^{m} I_k = \sum_{k=1}^{m} \tilde{I}_k \,,$$

with

$$\tilde{I}_k = \sum_{a,b,x,y} \alpha(a,b,x,y)\, p_{y+k-1\,\mathsf{mod(m}}(ab|xy)\,,$$

where $p_{y+k-1\,\mathsf{mod(m)}}$ denotes the conditional probability distribution between Alice and Bob number $y+k-1\,\mathsf{mod(m)}$.

If inequality (4.7) is violated, then $\tilde{I}_k > \beta$ for some k. Let us assume that this happens for $k=1$. For other value of k the proof follows similarly. $\tilde{I}_1$ then reads:

$$\tilde{I}_1 = \sum_{a,b,x,y} \alpha(a,b,x,y)\, p_y(ab|xy)$$

Notice that in the above expression, the terms involve only one measurement setting per Bob. That is, the terms that correspond to the j$^{\text{th}}$ Bob only involve measurement setting j for him. For these particular fixed inputs for the Bobs, let $p(ab_1 \ldots b_m|x\,1 \ldots m)$ denote the joint conditional probability distribution for the $m+1$ parties. In terms of it, $\tilde{I}_1$ takes the form

$$\tilde{I}_1 = \sum_{x,a} \sum_{b_1,\ldots,b_m} \alpha'(a,x,b_1,\ldots,b_m)\, p(ab_1 \ldots b_m|x\,1 \ldots m)\,, \tag{4.8}$$

where $\alpha'(a,x,b_1,\ldots,b_m) = \sum_y \alpha(a,b_y,x,y)$. Since the settings by the Bobs are fixed, by denoting $\vec{b} = (b_1,\ldots,b_m)$ we can rewrite the above expression as

$$\tilde{I}_1 = \sum_{x,a,\vec{b}} \alpha'(a,x,\vec{b})\, p(a,\vec{b}|x)\,.$$

Now, $p(a,\vec{b}|x) = p(a|\vec{b},x)\,p(\vec{b}|x) = p(a|\vec{b},x)\,p(\vec{b})$, since the No Signalling principle implies that $p(\vec{b}|x) = p(\vec{b})$. The crucial step is to realise that such a correlation $p(a,\vec{b}|x)$ can be realised by a LC model: a source prepares a system carrying the classical variable $\vec{b}$ with probability $p(\vec{b})$ and distributes it to the parties. Each Bob then outputs his corresponding b_j from $\vec{b}$, while Alice outputs a with probability $p(a|\vec{b},x)$. Since this is a LC model then $\tilde{I}_1 \leq \beta$, which contradicts our initial assumption that $\tilde{I}_1 > \beta$. Hence should the assumption hold, then Alice must signal to the Bobs.

Here I presented the argument for when $I(p)$ in eq. (4.6) corresponds to a bipartite Bell functional. However, a similar argument can be done when it corresponds to a multipartite Bell inequality, as shown in [59].

Finally, to illustrate this monogamy relations consider the case of the CHSH inequality. Using normalization constraints, this inequality can be rewritten as:

$$I(p) = \sum_{a,b,x,y} \delta_{a\oplus b=xy}\, p(ab|xy) \underset{\text{NCHV}}{\leq} 3 \underset{\text{Q}}{\leq} 2+\sqrt{2} \underset{\text{NS}}{\leq} 4\,.$$

Since here the number of settings is $m=2$, the construction provides a monogamy relation in the $(3,2,2)$ scenario. Hence, the expression results in

$$\sum_{a,b,x,y} \delta_{a\oplus b=xy}\, p(ab|xy) + \sum_{a,c,x,z} \delta_{a\oplus c=xz}\, p(ac|xz) \leq 6.$$

Beyond its fundamental relevance, the study of monogamy relations was boosted by their main role in quantum cryptography. Indeed, one can argue that the security proofs of quantum key distribution ultimately rely on the monogamy between Alice, Bob and Eve. These details will be discussed later on in the course.

4.5 From bipartite to multipartite scenarios

In this section I discuss how to handle in multipartite scenarios two topics preciously presented, namely the LO principle and almost quantum correlations.

4.5.1 Local Orthogonality: example

A "fun fact" about the LO principle is that, for bipartite scenarios, LO imposes the same constraints as the NS principle [46, 47] at the single-copy level. However, already for the simplest tripartite Bell scenario, namely $(3,2,2)$, the situation changes. Indeed, the Guess-Your-Neighbor's-Input (GYNI) inequality represents a family[1] of LO constraints that cannot be derived from the NS principle. For the $(3,2,2)$ scenario such an inequality reads:

$$p(000|000) + p(110|011) + p(011|101) + p(101|110) \leq 1\,.$$

Notice that $\{(000|000), (110|011), (011|101), (101|110)\}$ is a set of mutually orthogonal events.

For bipartite scenarios, then, nontrivial LO constraints are constructed by applying the principle at the many-copy level. To see how to do it, let us discuss now a particular example. Consider thus a CHSH scenario and let us see how to derive constraints by applying LO at the two-copy level. Since we are considering two copies of the CHSH scenario, then we have to study a four-partite scenario in the single-copy case. Hence, we need to first derive LO constraints at the single-copy level in a $(4,2,2)$ Bell scenario. Let us label the events in this scenario as $(a_1b_1a_2b_2|x_1y_1x_2y_2)$, where $(a_ib_i|x_iy_i)$ is an event for Alice and Bob in the i$^{\text{th}}$ CHSH scenario. Take now the following set of events:

$$S = \{(0000|0000), (1110|0011), (0011|0110), (1101|1011), (0111|1101)\}\,.$$

This set comprises events that are pairwise orthogonal. Hence, the LO principle imposes that

$$\sum_{(a_1b_1a_2b_2|x_1y_1x_2y_2)\in S} p(a_1b_1a_2b_2|x_1y_1x_2y_2) \leq 1\,.$$

Now, if $p(a_1b_1a_2b_2|x_1y_1x_2y_2)$ comes from the independent and parallel composition of two identical bipartite devices in the CHSH scenario in the way that the labels of the events suggest, it follows that

$$p(a_1b_1a_2b_2|x_1y_1x_2y_2) = p(a_1b_1|x_1y_1)p(a_2b_2|x_2y_2).$$

Hence,

$$\sum_{(a_1b_1a_2b_2|x_1y_1x_2y_2)\in S} p(a_1b_1|x_1y_1)p(a_2b_2|x_2y_2) \leq 1\,. \tag{4.9}$$

Eq. (4.9) is an example of a nontrivial LO constraint in the CHSH scenario that arises from considering two-copies of such. To see that it is non-trivial,

[1] A family constructed from an inequality that acts as a representative is generated by taking all the possible symmetries of such an inequality. These symmetries take into account relabeling of the input and output variables as well as permutation of the parties.

let us verify that PR-box correlations violate such condition. Indeed, for each $(a_1b_1a_2b_2|x_1y_1x_2y_2) \in S$, $p_{\mathrm{PR}}(a_1b_1|x_1y_1)p_{\mathrm{PR}}(a_2b_2|x_2y_2) = \frac{1}{2} \cdot \frac{1}{2}$. Hence,

$$\sum_{(a_1b_1a_2b_2|x_1y_1x_2y_2)\in S} p(a_1b_1|x_1y_1)p(a_2b_2|x_2y_2) = \frac{5}{4} > 1,$$

which violates LO.

4.5.2 Almost quantum correlations in multipartite Bell scenarios

The definition of almost quantum correlations can be generalised in a straightforward manner to mutipartite scenarios, although there is a little subtlety on the commutation relations of the operators. To keep the notation simple, consider a tripartite Bell scenario $(3, m, d)$ composed of Alice, Bob and Charlie. The events in this scenario are labeled by $(abc|xyz)$. Then, a conditional probability distribution $p(abc|xyz)$ is almost-quantum if there exists

- a Hilbert space $\mathcal{H}$,
- a projective measurement $\{\Pi_{a|x}\}_a$ in $\mathcal{H}$ for each x for Alice,
- a projective measurement $\{\Pi_{b|y}\}_b$ in $\mathcal{H}$ for each y for Bob,
- a projective measurement $\{\Pi_{c|z}\}_c$ in $\mathcal{H}$ for each z for Charlie,
- and a quantum state $\rho = |\psi\rangle\langle\psi|$ in $\mathcal{H}$

such that the statistics are recovered by them, i.e.

$$p(abc|xyz) = \mathrm{tr}\left\{\Pi_{a|x}\,\Pi_{b|y}\,\Pi_{c|z}\,\rho\right\}\,,$$

and the measurements operators for different parties satisfy the following commutation relations

$$\begin{aligned}\Pi_{a|x}\,\Pi_{b|y}\,\Pi_{c|z}\;|\psi\rangle &= \Pi_{c|z}\,\Pi_{a|x}\,\Pi_{b|y}\;|\psi\rangle\;,\\ &= \Pi_{b|y}\,\Pi_{c|z}\,\Pi_{a|x}\;|\psi\rangle\;,\\ &= \Pi_{b|y}\,\Pi_{a|x}\,\Pi_{c|z}\;|\psi\rangle\;,\\ &= \Pi_{c|z}\,\Pi_{b|y}\,\Pi_{a|x}\;|\psi\rangle\;,\\ &= \Pi_{a|x}\,\Pi_{c|z}\,\Pi_{b|y}\;|\psi\rangle\;.\end{aligned}$$

That is, any permutation of the operators $\Pi_{a|x}\,\Pi_{b|y}\,\Pi_{c|z}$ acting on the state $|\psi\rangle$ should yield the same result. This requirement is equivalent to "the operators should commute on the state" for bipartite scenarios, but much stronger than it for multipartite ones.

5 Contextuality

En esta unidad se introduce contextualidad mediante el teorema de Kochen-Specker y el cuadrado mágico de Mermin, y se presentan las desigualdades que evidencian contextualidad. Se menciona la conexión cualitativa con escenarios de Bell, y se reinterpretan los experimentos de nolocalidad como de contextualidad. Se presentan otras propuestas experimentales y se discuten sus limitaciones. Al final se presentan las bases del formalismo de Abramsky y Branderburger, y se desarrolla el enfoque de Cabello-Severini-Winter (CSW).

5.1 Kochen-Specker contextuality

In these past lectures, I introduced the phenomenon of Nonlocality. In a nutshell, a Bell experiment consists on distant parties, performing space-like separated actions on their share of a system. Bell's theorem then shows that no theory can make the same predictions as quantum theory, while jointly satisfying the properties of locality and realism[1], which are equivalent to the assumption of local causality.

In 1967, Kochen and Specker in a similar spirit derived another no-go theorem for quantum mechanics [60]. The Kochen-Specker (KS) theorem states that quantum theory is at variance with any attempt at assigning deterministic values to all observables in a way which would be consistent with the functional relationships between these observables predicted by quantum theory. This impossibility is generally known as contextuality, since it means that any potential 'hidden' predetermined value of an observable will necessarily have to depend on the context in which it is probed.

Similarly to nonlocality, contextuality tests usually deal with the statistics of measurement outcomes and expectations values of observables. However, in this case there are no space-like separated parties but just a single system under study. The idea then is, given a set of measurements/observables, check the compatibility of the observed outcome's statistics with classical or quantum models.

The Kochen-Specker theorem rules out a particular classical description of Nature; that is, one where the hidden variables models (a.k.a. ontic models) assign a deterministic value to each observable in the experiment, in a way that such a

[1] I am taking the free-will assumption for granted.

value does not depend on the context in which the observable is measured. In the particular case where the measurements in the scenario satisfy the compatibility relations of those in a Bell scenario, those classical models may equivalently be represented by indeterministic value assignments to the observables by the hidden variables: that is actually the equivalence between LD and LC models in Sec. 1. However, for general compatibility relations this is no longer the case. How to devise contextuality tests that rule out indeterministic non-contextual hidden variable models is beyond the scope of these lectures, but the keen student may find the answers in the work by Rob Spekkens [61]. For the purpose of this lecture, I will now present the KS theorem by a fully developed case study.

Consider the situation where we have five questions, $\{A_i\,,\,1 \leq i \leq 5\}$, each of which can give a yes/no answer. Moreover, assume that any two consecutive questions, $\{A_i, A_{i+1}\}$ where the sum in the sub-index is taken mod 5, can be asked simultaneously. But, why would we make such an assumption? I mean, if there are five questions why not ask them all together to begin with? This intuition, which is valid in the classical world, no longer holds in quantum mechanics. That is, there are sets of properties that one can measure on a system (i.e. questions that one can ask), which cannot all be performed together. In quantum mechanics, questions that cannot be asked simultaneously correspond to incompatible observables, and those that can correspond instead to jointly-measurable ones.

So going back to the example, we have five questions, and any two consecutive ones can be asked simultaneously. Such a situation is usually depicted as in Fig.5.2(a), where the vertices represent the questions and the edges join the compatible ones.

Kochen and Specker captured the intuition of "it should be possible to answer all the questions simultaneously" by what is now known as a deterministic noncontextual hidden variable (NCHV) model. Suppose we have a system, upon which we ask those questions. A NCHV model then assumes that the system comes equipped with a set of instructions, a.k.a. the hidden variables λ, that tell (deterministically) which questions give a 'yes' and which a 'no'. Or more generally, the system can be further equipped with a probability distribution $p(\lambda)$ on these hidden variables, and then when asked a question, answer with probability $p(\lambda)$ following the instructions given by λ.

Ultimately, we want to show that quantum mechanics predicts correlations that go beyond what NCHV models can explain. So similarly to Bell nonlocality, we will do it by studying inequalities.

Now back to the example. Since we ultimately want to study the statistics of the answers, we will assign a value $+1$ to the 'yes' answer, and a -1 to the 'no'.

In particular, we will focus on the figure of merit

$$K = \langle A_1\, A_2 \rangle + \langle A_2\, A_3 \rangle + \langle A_3\, A_4 \rangle + \langle A_4\, A_5 \rangle + \langle A_5\, A_1 \rangle \,.$$

Now we want to know the minimum value that such an expression can take when the statistics of the system that is being probed can be explained by a NCHV model. That is, we want to compute

$$\begin{aligned} \min \quad & a_1 a_2 + a_2 a_3 + a_3 a_4 + a_4 a_5 + a_5 a_1 \\ \text{st} \quad & a_i = \pm 1 \quad \forall\, i \,. \end{aligned}$$

Note that, similarly to nonlocality, the minimisation is being performed only over the deterministic assignments λ rather than over the whole convex space that they define (i.e. all the possible NCHV models). The reason for this is that any model which is not an extremal deterministic one can only yield a value of K larger than the minimum one of the deterministic points in its decomposition. A straightforward calculation then gives

$$\langle A_1\, A_2 \rangle + \langle A_2\, A_3 \rangle + \langle A_3\, A_4 \rangle + \langle A_4\, A_5 \rangle + \langle A_5\, A_1 \rangle \underset{\text{NCHV}}{\geq} -3 \,. \qquad (5.1)$$

Indeed, one can see that if any four of the terms $a_i a_{i+1}$ are set to be -1, the remaining one automatically yields a $+1$ value, rendering their sum -3. Eq. (5.1) is usually referred to as the Klyachko-Can-Binicioğlu-Shumovsky (KCBS) inequality [62].

To see that quantum theory violates the KCBS inequality, consider the case of a qutrit prepared on the pure state $|\psi\rangle = [0, 0, 1]^T$. In addition, choose the following dichotomic observables:

$$A_1 = \frac{2}{N}\begin{bmatrix}\cos\left(\frac{4\pi}{5}\right)\\ \sin\left(\frac{4\pi}{5}\right)\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix}\cdot\left[\cos\left(\frac{4\pi}{5}\right),\sin\left(\frac{4\pi}{5}\right),\sqrt{\cos\left(\frac{\pi}{5}\right)}\right]-\mathbb{1}\,,$$
$$A_2 = \frac{2}{N}\begin{bmatrix}\cos\left(\frac{2\pi}{5}\right)\\ -\sin\left(\frac{2\pi}{5}\right)\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix}\cdot\left[\cos\left(\frac{2\pi}{5}\right),-\sin\left(\frac{2\pi}{5}\right),\sqrt{\cos\left(\frac{\pi}{5}\right)}\right]-\mathbb{1}\,,$$
$$A_3 = \frac{2}{N}\begin{bmatrix}\cos\left(\frac{2\pi}{5}\right)\\ \sin\left(\frac{2\pi}{5}\right)\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix}\cdot\left[\cos\left(\frac{2\pi}{5}\right),\sin\left(\frac{2\pi}{5}\right),\sqrt{\cos\left(\frac{\pi}{5}\right)}\right]-\mathbb{1}\,,$$
$$A_4 = \frac{2}{N}\begin{bmatrix}\cos\left(\frac{4\pi}{5}\right)\\ -\sin\left(\frac{4\pi}{5}\right)\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix}\cdot\left[\cos\left(\frac{4\pi}{5}\right),-\sin\left(\frac{4\pi}{5}\right),\sqrt{\cos\left(\frac{\pi}{5}\right)}\right]-\mathbb{1}\,,$$
$$A_5 = \frac{2}{N}\begin{bmatrix}1\\ 0\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix}\cdot\left[1,0,\sqrt{\cos\left(\frac{\pi}{5}\right)}\right]-\mathbb{1}\,,$$

where the normalisation constant is $N = 1 + \cos\left(\frac{\pi}{5}\right)$. One can check that these observables are compatible according to our premises, since (i) $[A_i, A_j] = 0$ only if $i = j \pm 1$, and (ii) the observables are constructed from projective measurements, hence commutativity is necessary and sufficient for joint-measurability. Notice that these observables are expressed as $A_k = 2\,|v_k\rangle\,\langle v_k| - \mathbb{1}$, where $|v_k\rangle$ are some normalised vectors in $\mathbb{R}^3$. A pictorial representation of these $\{|v_k\rangle\}_k$ together with the vector $|\psi\rangle$ is given in fig. 5.1.

A straightforward calculation gives a quantum value for KCBS of $K = 5-4\sqrt{5} \sim -3.94427$, hence violating ineq. (5.1).

The KS theorem hence tells us that quantum theory is incompatible with a world where the systems can only be prepared on states which are mixtures of those that only assign deterministic values to measured observables.

Inequalities such as that of KCBS given in (5.1) are usually called contextuality inequalities. Contextuality proofs that rely on inequality violation by a set of observables and a particular quantum state, such as the one discussed above, are referred to as *state dependent* proofs of contextuality.

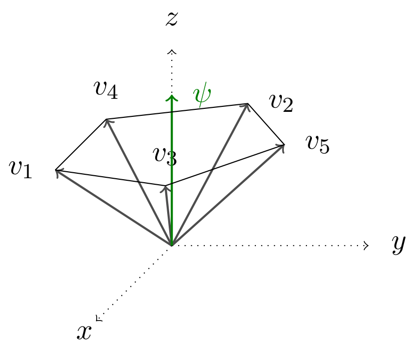


Figure 5.1: The caption

5.2 State-independent contextuality

There is an even stronger form of contextuality than the one discussed before. That is, one where the contradiction with NCHV models arises merely by the choice of observables, i.e. it holds for any choice of the state of the system. This type of contextuality proofs are called *state independent*.

One example of such is the so called *Peres-Mermin square* [63, 64], and applies to two-qubit systems. Consider the following set of nine observables $\{A, B, C, a, b, c, \alpha, \beta, \gamma\}$:

$A = \sigma_z \otimes \mathbb{1}$	$B = \mathbb{1} \otimes \sigma_z$	$C = \sigma_z \otimes \sigma_z$
$a = \mathbb{1} \otimes \sigma_x$	$b = \sigma_x \otimes \mathbb{1}$	$c = \sigma_x \otimes \sigma_x$
$\alpha = \sigma_z \otimes \sigma_x$	$\beta = \sigma_x \otimes \sigma_z$	$\gamma = \sigma_y \otimes \sigma_y$

where σ_k, $k \in \{x, y, z\}$, are the Pauli matrices. Here, the observables in any row or column commute and are therefore compatible. Note that these compatibility relations do not arise from space-like separation constraints, and hence this cannot be thought of as a Bell experiment even though it consists of two-qubits.

In addition, the product of the observables in any row or column equals $\mathbb{1}$, but for the last column where it equals $-\mathbb{1}$. Hence, the following holds for *any* state of the system:

$$\langle ABC \rangle + \langle abc \rangle + \langle \alpha\beta\gamma \rangle + \langle Aa\alpha \rangle + \langle Bb\beta \rangle - \langle Cc\gamma \rangle = 6\,. \tag{5.2}$$

On the other hand, we now want to know what is the maximum value that NCHV models can yield for the left-hand side of eq. (5.2). To compute such a maximum then one needs to optimise over all the deterministic assignment of values ± 1 to the observables $\{A, B, C, a, b, c, \alpha, \beta, \gamma\}$. A straightforward calculation gives

$$\langle ABC\rangle + \langle abc\rangle + \langle \alpha\beta\gamma\rangle + \langle Aa\alpha\rangle + \langle Bb\beta\rangle - \langle Cc\gamma\rangle \underset{\text{NCHV}}{\leq} 4\,.$$

We see then that quantum theory violates this inequality for the particular chosen observables and for *all* quantum states.

These state-independent proofs of contextuality show how any quantum state reveals non-classical properties if the measurements are chosen appropriately.

5.3 Inequalities from hypergraphs: CSW

Similarly to nonlocality, when exploring contextuality the use of inequalities comes in quite handy. Hence, particular effort has been devoted to deriving contextuality inequalities. Of particular relevance is the work done by Cabello, Severini and Winter (CSW), who defined families of inequalities associated to each compatibility structure of a set of events (a graph) [65]. They hence used tools from graph-theory to study the phenomenon of contextuality and its potential realisation beyond what quantum mechanics allows. In this section I will review their approach.

So let us first start from the very beginning. What is a set of "events"? An *event* basically tells you the question/context that is measured and the values that have been obtained for each measured observable. For instance, in the KCBS setup discussed before, the set of all events involved in the experiment is given by

$$\{(a_i, a_{i+1}|A_i, A_{i+1})\,|\,a_i, a_{i+1} = \pm 1\,,\, 1 \leq i \leq 5\},$$

where the sum is taken mod 5. Hence every description of a contextuality scenario in terms of its sets of questions and their compatibility relations will give rise to a set of events.

In order to illustrate the CSW method, let me review as a first case study an equivalent formulation of the KCBS scenario[2]. Similarly to before, consider five questions $\{P_i\,,\, 1 \leq i \leq 5\}$ with yes/no answers. We will assume that this set of questions has the following properties:

(i) For each i, P_i and P_{i+1} (where the sum is taken mod 5) are compatible. I.e., they can be jointly asked without mutual disturbance, so, when the questions are repeated, the same answers are obtained.

(ii) For each i, P_i and P_{i+1} are exclusive. That is, they can't be both simultaneously answered with 'yes'.

[2]Strictly speaking, these two formulations are equivalent only for quantum theories. For general theories some extra assumptions need to be considered in order to make the claim. Nowadays most people agree that the logic behind the first formulation of KCBS is more appropriate to define contextuality scenarios in general theories.

Now, what is the maximum number of 'yes' that one can obtain when these five questions are asked to a physical system? If the system's statistics is consistent with a NCHV model, then the answer is two. This is a direct consequence of the exclusivity property. If we associate outcome 'yes' to 1 and 'no' to 0, this fact can be mathematically stated as

$$\sum_{i=1}^{5} \langle P_i \rangle \underset{\text{NCHV}}{\leq} 2\,. \tag{5.3}$$

Now, how is this related to the previous KCBS scenario that we discussed? Here, we focused on the answers given by single questions, where before we worked with the correlations that these outcomes have when we ask compatible ones together. Notice moreover that before, each question A_i was ultimately thought of as a dichotomic observable with possible values -1 for 'no' and $+1$ for 'yes'. This way, when both questions give the same answer the product of the results of the observables is $+1$, but when the answers are different then the product of the results of the observables is -1.

This suggests the following relationship between the random variables A_i and $P_i:$ $A_i = 2P_i - 1$. When we ask question i and obtain a 'yes', P_i is assigned a 1 and A_i a 1; when the answer is 'no' P_i is assigned a 0 and A_i a -1. The next step is to see that, eq. (5.3) together with the exclusivity condition (ii) imply eq. (5.1). First notice that

$$A_i A_{i+i} = 4P_i P_{i+1} - 2P_i - 2P_{i+1} + 1\,, \quad \Rightarrow \quad \langle A_i A_{i+i} \rangle = -2\langle P_i \rangle - 2\langle P_{i+1} \rangle + 1\,,$$

since (ii) implies $\langle P_i P_{i+1} \rangle = p(\text{yes}, \text{yes}|P_i P_{i+1}) = 0$. Now it's just a matter of some algebra to go from (5.3) to (5.1).

As a remark, before we were able to derive eq. (5.1) without assuming any exclusivity constraints (only compatibility ones). This is because the particular figure of merit that we were optimising over implements a penalty for violating exclusiveness.

So, back to the example . Our starting point is the scenario where we have five questions $\{P_i\}$ that satisfy (i) and (ii). Is there a systematic way to find the inequality (5.3) without having to resort to divine inspiration? The answer is 'yes', and was given by CSW in terms of graph theoretical tools.

The first step of the approach is to construct a graph where the vertices are the events of the scenario. Here, since we ask single questions to the system, these will be given by $\{(0|P_i), (1|P_i)\}_i$. Now we are interested in inequalities which are linear functionals of the expectation values of P_i, where $\langle P_i \rangle = p(1|P_i)$. Hence, since the events of the form $(0|P_i)$ are rendered irrelevant, among the full list of events we will only keep those of the form $(1|P_i)$. Now CSW defines the *exclusivity* *graph* G for these events as that whose vertices are given by $V(G) = \{(1|P_i)\}$, and two events share an edge if they are exclusive. The exclusivity condition imposed via (ii) then tells that G takes the form of a pentagon, as in Fig. 5.2(b). In general scenarios, a natural notion of exclusiveness can be defined without having to impose any extra constraints. Later on we will discuss this and an example.

In this KCBS scenario, the type of inequalities that we are looking for then have the form:

$$\sum_{i=1}^{5} \alpha_i \langle P_i \rangle \underset{\text{NCHV}}{\leq} \alpha \,, \tag{5.4}$$

where the $\{\alpha_i\}$ and α are real numbers. The question is then, given $\{\alpha_i\}$ how to find α. Graph theory gives us the answer straight away: α is the *weighted independence number* of G, when equipped with the weights α_i. That is, first associate to each vertex i the weight $w(i) = \alpha_i$, and then compute

$$\begin{aligned} \alpha = \max_I \sum_{v \in I} w(v) & \\ & \text{st} \quad I \quad \text{is an independent set of vertices.} \end{aligned}$$

Recall that an independent set I is a set of vertices such that no pair in in it share an edge. When the coefficients are $\alpha_i = 1$ for all i, we get $\alpha = 2$ and hence recover eq. (5.3).

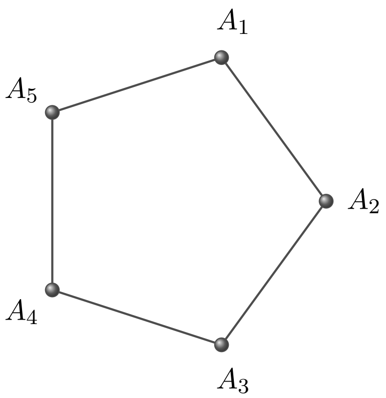
Note that the same reasoning can actually be done in a more general way starting from the exclusivity graph defined from the full set of events[3] $\{(0|P_i), (1|P_i)\}_i$, and now studying inequalities of the form

$$\sum_{i=1}^{5} \alpha_i \, p(1|P_i) + \beta_i \, p(0|P_i) \underset{\text{NCHV}}{\leq} \alpha \,. \tag{5.5}$$

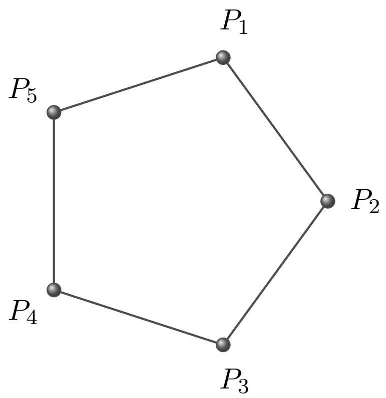
Now in this new exclusivity graph, depicted in Fig. 5.2(c), two vertices will share an edge unless they are a pair $\{(1|P_i), (1|P_{i+1})\}$ for any i. When we set $\alpha_i = 1$ and $\beta_i = 0$ for all i we recover $\alpha = 2$ and eq. (5.3).

So, what is the new insight that the method by CSW brings into the problem? I mean, computationally finding the NCHV bound via an optimisation over deterministic NCHV assignments is as complex as computing the weighted independence number. That is, so far it might seem like a rephrasing of the problem. But like someone once said "with new perspective comes new insight". Now that we have a problem phrased in terms of graph theory, we can use graph theory to explore it.

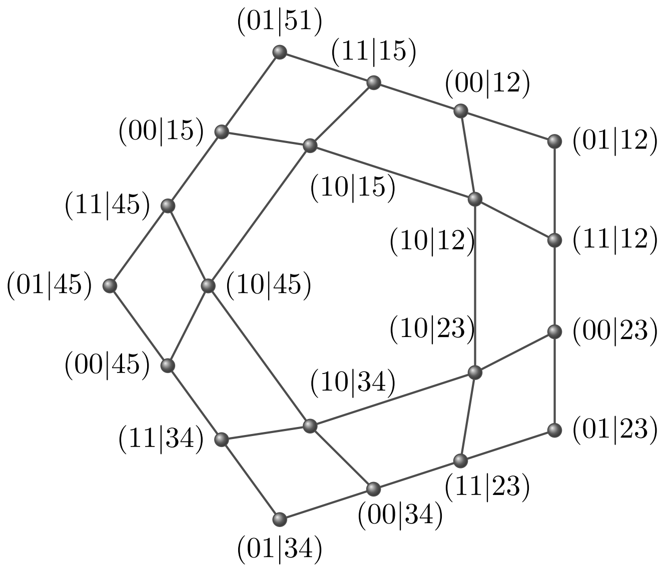
[3]The other way though lets us get to the nice pentagon for the KCBS inequality straight away.



(a) KCBS scenario: first formulation.



(b) KCBS scenario: second formulation.



(c) Exclusivity graph for the KCBS scenario.

Figure 5.2: KCBS scenario: (a) compatibility graph for the first formulation of KCBS, (b) exclusivity graph for the second formulation of KCBS.

On the one hand, graphs and their independence numbers have been vastly studied. We hence can use known examples of graphs to generate potentially interesting contextuality scenarios.

On the other hand, the most important contribution comes when studying the values that the figure of merit can achieve when the statistics are asked to be compatible with the predictions of quantum theory. In other words, what is the maximum violation that the inequality can have when the questions are asked on a quantum system. The answer to this is the following: the *weighted Lovász number* of G. In a nutshell, the Lovász number of a graph G is given by

$$\vartheta(G,w) = \sum_{v\in V} w(v)\,|\langle\phi_v|\Psi\rangle|^2\,,$$

where $|\Psi\rangle$ and $\{|\phi_v\rangle\}_v$ are unit vectors, and $\langle\phi_v|\phi_u\rangle = 0$ whenever v and u share an edge in G. Actually, to compute the Lovász number, an optimisation of such sum over all possible $|\Psi\rangle$ and $\{|\phi_v\rangle\}_v$ with those properties is performed.

The beauty of this is that now we can think of $|\Phi\rangle$ as the state of a quantum system, and of $|\phi_v\rangle\langle\phi_v|$ as the projectors associated to the answer v. For instance, let us go back to our example of the KCBS scenario. There, the vectors that achieve the Lovász number are given by

$$|\phi_{1,4}\rangle = \frac{1}{\sqrt{N}}\begin{bmatrix}\cos\left(\frac{4\pi}{5}\right)\\ \pm\sin\left(\frac{4\pi}{5}\right)\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix},\ |\phi_{2,3}\rangle = \frac{1}{\sqrt{N}}\begin{bmatrix}\cos\left(\frac{2\pi}{5}\right)\\ \mp\sin\left(\frac{2\pi}{5}\right)\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix},$$
$$|\phi_5\rangle = \frac{1}{\sqrt{N}}\begin{bmatrix}1\\ 0\\ \sqrt{\cos\left(\frac{\pi}{5}\right)}\end{bmatrix},\ |\Psi\rangle = \begin{bmatrix}0\\0\\1\end{bmatrix},$$

where the normalisation constant is $N = 1+\cos\left(\frac{\pi}{5}\right)$. These vectors are precisely the ones depicted in fig. 5.1. In this case, the Lovász number is $\vartheta(KCBS) = \sqrt{5}\sim 2.236$ which is larger than the NCHV bound of 2.

A couple remarks are in order. First, computing the Lováz number of a graph is a semidefinite program (SDP), which are considered 'easy' to solve. Second, any graph G with the property that $\alpha(G) < \vartheta(G)$ exhibits a gap between what is classically noncontextually achievable and what quantum theory reaches, which can be witnessed in experiments with an appropriate set of projectors and an appropriate state. Hence, every such graph can provide a proof of the KS theorem. The same extends to the case where weights are added to the terms in the inequality. Hence, CSW provide a way to identify potentially useful scenarios for contextuality experiments[4].

Let us see another example: a bipartite Bell scenario with two dichotomic inputs per party. Actually, I still haven't discussed how Bell nonlocality fits within contextuality experiments, and I will make that precise next lecture. But for the time being, let me discuss this example. The compatibility structure for the Bell setup is given as in Fig.5.3(a), where the compatible observables are $\{A_i, B_j\}$ for any choice of i, j. This compatibility is guaranteed by the space-like separation between the parties. The full set of events is given by $\{(ab|xy) \,:\, a, b, x, y = 0, 1\}$. Now the question is how to construct the exclusivity graph. This is done by importing the orthogonality notion defined by what was later known as the *local orthogonality principle*:

Definition 5.1. *Two events are orthogonal (a.k.a. exclusive) if 'both cannot be simultaneously true': i.e. there exists an observable that appears in both contexts but the outcomes obtained for it in each event differ. For instance, in the KCBS scenario,* $(1, 1|A_1, A_2)$ *is orthogonal to* $(-1, 1|A_2, A_3)$*. Moreover, the orthogonality relation between* $(1, 1|A_1, A_2)$ *and* $(-1, -1|A_3, A_4)$ *cannot be determined since the events do not have an observable in common.*

With this, the orthogonality graph G of the CHSH scenario can be depicted as in Fig. 5.3(b). Now, the CHSH Bell inequality can be rewritten as

$$\sum_{\substack{ab\\a=b}} p(ab|00) + \sum_{\substack{ab\\a=b}} p(ab|10) + \sum_{\substack{ab\\a=b}} p(ab|01) + \sum_{\substack{ab\\a\neq b}} p(ab|11) \underset{\text{NCHV}}{\leq} 3\,. \tag{5.6}$$

So let us now apply the graph-theoretic tools to the exclusivity graph G equipped with the weights $w(ab|xy) = \delta_{a\oplus b=xy}$. Given that only eight events have a nonzero coefficient, the effective graph that we need to study is the eight-vertex circulant $(1, 4)$ graph $Ci_8(1, 4)$, depicted in Fig. 5.3(c). We hence find that $\alpha(G, w) = 3$ and $\vartheta(G, w) = 2+\sqrt{2}$, which are indeed the classical bound and Tsirelson's bound for eq. (5.6).

Before ending this session, two more things need to be discussed. First, about correlations beyond what quantum theory allows. Second, about some technicalities and limitations of CSW when tackling Bell scenarios.

Let us begin by the former. As was mentioned in previous lectures, quantum theory is our current most accurate description of Nature, but it is not certain that it is the ultimate theory that describes it. So people like exploring up to what extent we can violate inequalities beyond Tsirelson's bounds by still sensible correlations. The CSW approach hence gives us tools to compute the maximum value that

[4] I say 'potentially' because the actual state and measurements required to witness the gap might be experimentally challenging, which ultimately reduces their applicability.

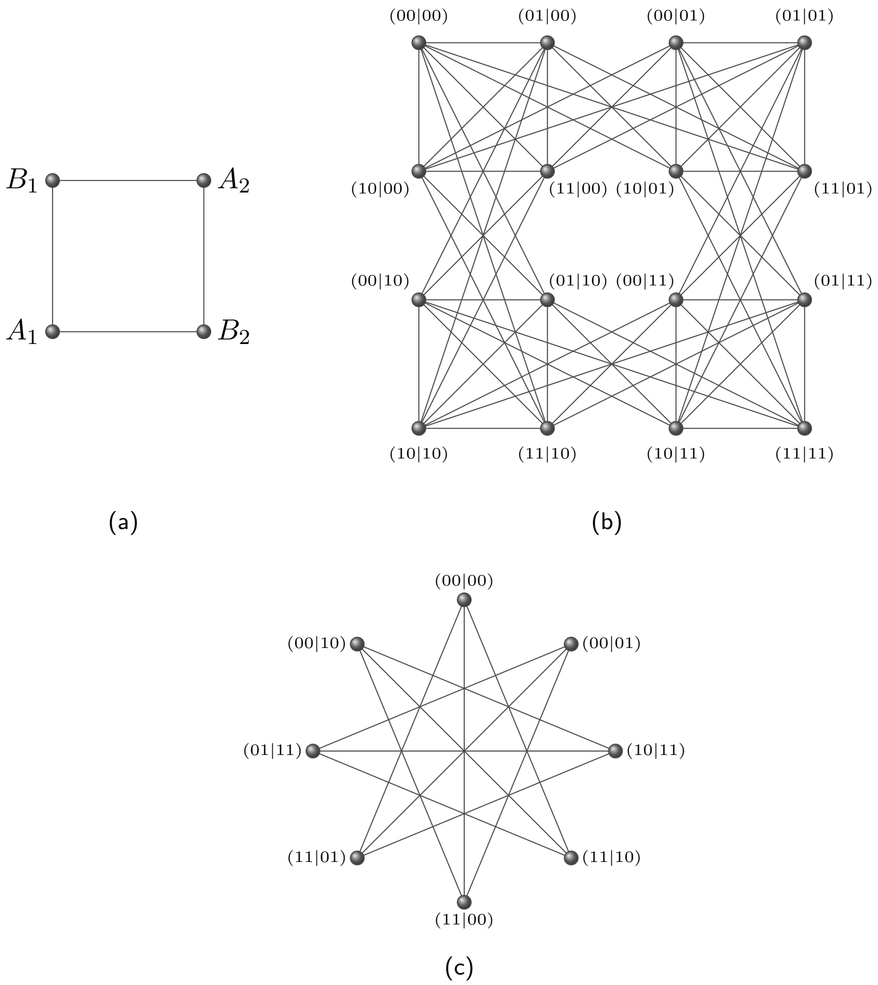


Figure 5.3: CHSH scenario: (a) compatibility graph for the dichotomic observables, (b) exclusivity graph of the full scenario, (c) orthogonality graph corresponding to the CHSH inequality (5.6).

correlations compatible with the Consistent Exclusivity (CE) principle achieve. CE imposes that "the sum of the probabilities of pairwise mutually exclusive events cannot exceed 1". Hence, to obtain the maximum violation of an inequality we need to maximise the figure of merit with the constraint that $\sum_{v\in C} p(v) \leq 1$ for every clique[5] C in G. CSW then shows that this value is given by the *weighted fractional packing number* α^* of (G, w). For the second formulation of KCBS, $\alpha^* = \frac{5}{2}$ and for the CHSH scenario $\alpha^* = 4$, hence both experiments could potentially serve to test correlations beyond quantum mechanics. Moreover, in the CHSH case that bound coincides with the No Signalling bound. This is consistent with the fact that for bipartite Bell scenarios CE is equivalent to the No Signalling principle.

The final remark is about issues between CSW and Bell scenarios. Long story short, when applying their technique directly to Bell inequalities, the Lovász number ultimately gives an upper bound to Tsirelson's bound, which in the CHSH case happens to be tight. This is because there are two main aspects that we need to further consider in Bell scenarios:

(i) The probabilities need to be properly normalised. I.e. $\sum_{v\in e} |\langle\phi_v|\Psi\rangle|^2 = 1$ for every set of events e that defines a complete measurement. In general, computing the maximum via the Lovász number alone only guarantees $\sum_{v\in e} |\langle\phi_v|\Psi\rangle|^2 \leq 1$. Since this is a linear constraint, optimising over $|\Psi\rangle$ and $\{|\phi_v\rangle\}$ is still efficient to solve.

(ii) The projectors need to define a von Neumann measurement. I.e. $\sum_{v\in e} |\phi_v\rangle\langle\phi_v| = \mathbb{1}$ for every set of events e that defines a complete measurement. Imposing this constraint is not trivial, and how to perform this optimisation in an efficient way is still an open problem.

CSW hence supplements the calculations of quantum-Bell maxima with condition (i) to obtain better upper bounds to Tsirelson's bound. Take for instance the case of the I_{3322} Bell inequality, where Tsirelson's bound is known to lie below 0.2508755. Here, a computation of the Lovász number gives a value of 0.4114, while when complemented with (i) yields 0.25147. Both results are nevertheless strictly larger than Tsirelson's bound.

5.4 Compatible observables scenarios: the sheaf theoretic approach

Abramsky and Branderburger were the first to provide a framework where to study nonlocality and contextuality in a unified manner. Their framework is mathematically based on sheaf-theory, the study of which goes beyond the scope of these

[5] A clique in a graph is a fully connected subgraph.

lectures. In this section, however, I will briefly present their definitions, tu set up the language that I will use to formally refer to compatible-observables scenarios.

In [66], a contextuality scenario is referred to as a *marginal scenario* and defined as follows:

Definition 5.2. *A* ***marginal scenario*** $(X, O, \mathcal{M})$ *is a finite set* X*, the elements of which we call* ***observables****, together with a finite set* O *of outcomes and a* ***measurement cover*** $\mathcal{M}$*, which is a family of subsets* $\mathcal{M} \subseteq 2^X$ *such that*

1. *every element of* X *occurs in some* C*, i.e.* $\bigcup_{C\in\mathcal{M}} C = X$*.*
2. $\mathcal{M}$ *is an anti-chain: for any* $C, C' \in \mathcal{M}$*, if* $C \subseteq C'$*, then* $C = C'$*.*

The $C \in \mathcal{M}$ *are called* ***measurement contexts****.*

In other words, a contextuality scenario is hence given by a set of observables, a set of outcomes that those observables may yield, and a list of which maximal sets of observables are jointly measurable. As noted in [66], it is not a substantial restriction to assume that all observables take values in the same set of outcomes O.

These scenarios, up to the information on O, may be depicted by a hypergraph, whose vertices are given by the observables, and the hyperedges by each element of the measurement cover. See Fig. 5.3(a) for an example, which shows the CHSH scenario as a marginal scenario with observables A_1, A_2, B_1, B_2 where the four pairs

$$\{A_1, B_1\}, \quad \{A_1, B_2\}, \quad \{A_2, B_1\}, \quad \{A_2, B_2\}$$

are jointly measurable, but no other pairs or triples of observables are jointly measurable. In particular, these four pairs also are the maximal sets of jointly measurable observables and thereby form the measurement cover

$$\mathcal{M} = \{\{A_1, B_1\}, \{A_1, B_2\}, \{A_2, B_1\}, \{A_2, B_2\}\}.$$

Since each hyperedge consists of two vertices, we have depicted them as just edges in the figure.

The outcomes statistics that are studied in these experiments are called *empirical models*, and here I will denote them by P. For each measurement context $C \in \mathcal{M}$, P defines a probability distribution P_C over O^C, such that the *sheaf condition* holds:

$$P_{C|C\cap C'} = P_{C'|C\cap C'} \quad \forall C, C' \in \mathcal{M}, \tag{5.7}$$

where $P_{C|C\cap C'}$ stands for the marginal distribution of P_C associated to the observables in $C \cap C'$. For an assignment of outcomes $s \in O^C$, the probability $P_C(s)$ is to be thought of as the probability of obtaining the joint outcome s when jointly measuring all observables in C. The sheaf condition is a generalization of the no-signaling condition, also referred to as *no-disturbance*.

Similar to the case of Bell nonlocality, depending on which model of Nature we use, different sets of empirical models arise. Generic models that merely comply with the no-disturbance condition are sometimes called *general empirical models*. *Classical empirical models* are those that arise as convex combinations of deterministic non-contextual value assignments to the observables, just as discussed earlier on in this lecture. Finally, *quantum empirical models* are whose where each observable can be identified with a complete projective measurement, and there exists a quantum system such that when those measurements are performed on it the experimental statistics are recovered.

5.5 Contextuality bundles

The compatible-observable framework discussed in the previous section gives rise, among other things, to a geometrical picture of contextuality proofs and a hierarchy of strengths of contextuality [67]. In this section I will briefly review this graphical results with some examples.

The starting point of the technique is an empirical model. From it, a possibility table is constructed by assigning value 1 to each event that happens with non-zero probability, and 0 otherwise. This possibility table is then depicted in a useful manner. Finally, from this plot we decide if the model is contextual, and *how strong* its contextuality is.

For the purpose of this lecture, let us explore the study case of a CHSH Bell scenario, and three empirical models: a deterministic classical one, a quantum one, and a postquantum nonsignalling one.

Let us start from the deterministic classical model given by $p_c(ab|xy) = (\delta_{0,a}\delta_{0,x}+\delta_{1,a}\delta_{1,x})\,\delta_{0,b}$. This model deterministically outputs 0 for Alice when she measures A_1, 1 when she measures A_2, and always 0 for Bob regardless of his measurement choice. Table 5.4(a) represents p_c. Now we construct the behaviour's *possibility table*. This is given by a table similar to the probability one, but where now the entry $(ab|xy)$ is 1 if $p(ab|xy) \neq 0$, and 0 otherwise. For the particular case of p_c, since it's deterministic, its possibility table coincides with its probability one.

So now the idea is to turn the possibility table into a 3-D diagram. First, along the $x-y$ plane, we will draw the compatibility graph of the scenario. Then we will copy such a graph into the $x-y$ plane at $z=1$. Finally, we will draw those edges that appear in the possibility table by interpreting the node in the z plane for the observable as outcome z for that observable. See Fig. 5.7(a) for the

	$A_1\,B_1$	$A_2\,B_1$	$A_1\,B_2$	$A_2\,B_2$
00	1	0	1	0
10	0	1	0	1
01	0	0	0	0
11	0	0	0	0

(a) Probability table for p_c.

	$A_1\,B_1$	$A_2\,B_1$	$A_1\,B_2$	$A_2\,B_2$
00	1	0	1	0
10	0	1	0	1
01	0	0	0	0
11	0	0	0	0

(b) Possibility table for p_c.

Figure 5.4: Probability and possibility tables for the classical deterministic behaviour p_c.

diagram corresponding to p_c. The segment joining two possible outcomes for two compatible observables is called a *local section*. For instance, the green segment joining the outcomes 0 for observables A_1 and B_1 in Fig. 5.7(a) is a local section. A collection of local sections one from each context is a *global section*. For instance, the green polygon in a possible global section for the behaviour p_c. This particular classical model has only one possible local section per context, and their collection gives a well defined global sections. For a given local section, if there exists a global one that contains it, we say that *the local section can be extended to a global one*. In the example of p_c, every local section can be extended to a global one (in the case, The global one).

Now let us study a more challenging example, given by the empirical model p_q given in [68]. The possibility table of such a behaviour is given by Table 5.5(a). Now let us look at the corresponding bundle diagram for p_q, which is depicted in Fig. 5.7(b). There, the local sections depicted in green are an example of a global section allowed by the model. The red local section, however, happens not to be extendible to any global section. Indeed, the only choices of compatible local sections for the contexts (A_2B_1) and (A_1B_2) are those depicted in blue, but then no choice of local section for the context (A_2B_2) satisfies all the other local constraints. The existence of such a local section with no global extension is a signature of nonclassicality. In other words, the behaviour p_q is contextual in the KS sense.

Whenever a behaviour has the property that "there exists a local section that has no global extension", it is *logically contextual*. Their contextuality can be signalled by the violation of a logical contextuality inequality. Not every contextual model is logically contextual.

Finally, consider the case of PR box correlations. Their probability and possibility tables are given in 5.6(a) and 5.6(b) respectively. The corresponding bundle diagram is depicted in Fig. 5.7(c). Direct inspection shows that in this case no local section can be extended to a global one. Whenever this happens, as is the

	$A_1\,B_1$	$A_2\,B_1$	$A_1\,B_2$	$A_2\,B_2$
00	1	0	0	1
10	1	1	1	1
01	1	1	1	1
11	1	1	1	0

(a) Possibility table for p_q.

Figure 5.5: Probability and possibility tables for the quantum behaviour p_q.

	$A_1\,B_1$	$A_2\,B_1$	$A_1\,B_2$	$A_2\,B_2$
00	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
10	0	0	0	$\frac{1}{2}$
01	0	0	0	$\frac{1}{2}$
11	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0

(a) Probability table for p_{PR}.

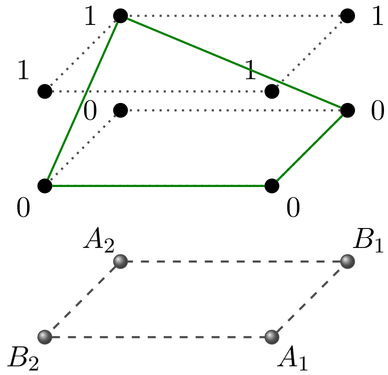
	$A_1\,B_1$	$A_2\,B_1$	$A_1\,B_2$	$A_2\,B_2$
00	1	1	1	0
10	0	0	0	1
01	0	0	0	1
11	1	1	1	0

(b) Possibility table for p_{PR}.

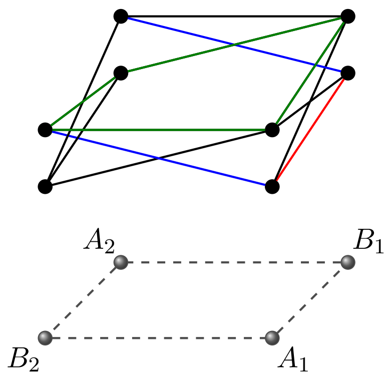
Figure 5.6: Probability and possibility tables for the postquantum nonsigmalling behaviour p_{PR}: a PR box.

case for p_{PR}, the behaviour is said to be *strongly contextual*. These models are a subset of the logically contextual ones.

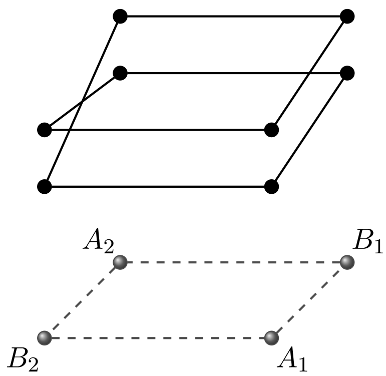
This new way of looking at empirical models in contextuality scenarios provides us with a geometrical picture where to study their degree of contextuality (if any) with topological tools. The particulars of this goes beyond the scope of these lectures, and the keen student is welcome to check [67].



(a) Bundle for the deterministic classical model p_c



(b) Bundle for the quantum model p_q.



(c) Bundle for the nonsignalling model p_{PR}.

Figure 5.7: Bundle diagrams for a (a) classical, (b) quantum, and (c) non-signalling empirical models discussed in the text.

6 A hypergraph framework for contextuality

En esta unidad se presenta el formalismo de hypergrafos de Acín et al. Se introducen el concepto de modelos probabilísticos, entre ellos los clásicos (no-contextuales), cuánticos y generales, los politopos y conjuntos de correlaciones y la interpretación geométrica de las desigualdades de contextualidad. Se desarrolla luego la correcta representación de escenarios de Bell en el formalismo de contextualidad, y se presenta el problema de caracterizar el conjunto de modelos probabilísticos cuánticos. Luego se introduce una jerarquía de relajaciones del conjunto cuántico y los modelos probabilísticos casi-cuánticos. Finalmente se hace la conexión con teoría de grafos, mostrando como diferentes graph-invariants caracterizan ciertos conjuntos de modelos probabilísticos.

6.1 Scenarios with operational equivalences

Acín, Fritz, Leverrier and Sainz (AFLS) [69] developed a dual approach to that of CSW [65]. In AFLS, the sets of classical, quantum, and general probabilistic models are themselves the primary objects, rather than their maximum value for contextuality inequalities. Although in principle both approaches are equivalent by duality, working directly with the behaviours is a more natural thing to do, since the actual quantities gathered from an experiment are outcome probabilities rather than coefficients of some inequality, Moreover, satisfaction of a predetermined inequality is sufficient, but not necessary, for the measured statistics to arise from a classical or quantum model. The main advantage of AFLS is that it explicitly takes into account the normalization of probabilities from the very beginning, and allows the inclusion of Bell scenarios within contextuality ones in the proper manner.

So let us begin by stating precisely what AFLS considers as a contextuality scenario. In a nutshell, a scenario is defined by

(i) a set of events,

(ii) a set of complete measurements,

(iii) operational relations between the events.

Similarly to before, the set of events is given by the collection of measurement outcomes $\{v = a|x\}$, i.e. the answers of the questions that can be asked to the system in the particular experimental setup. Each *event* represents an outcome obtained from the system after it receives some input or "measurement choice". However, now the measurements are considered complete, in the sense that whenever one is performed an outcome is always obtained, and that correspond to an event in the set.

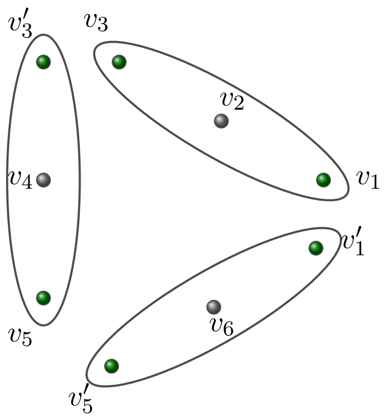
The other main difference to CSW are the operational equivalences that relate events, and which are not assumed to arise via considerations of joint measurability of observables. The list of operational equivalences tell you which events can be thought of as equivalent. Operationally, the idea goes as follows. Suppose that I have two different measurements x and x', and that whenever I perform them on a system I obtain the events $a|x$ and $a'|x'$ with the same probability, independently of the state the system is in. From an operational perspective, there is no distinction between $a|x$ and $a'|x'$, and hence are rendered equivalent. For example, consider the case of quantum theory, and take the following projective measurements: $\{\Pi_0, \Pi_1, \Pi_2\}$ and $\{\Pi_3, \Pi_0, \Pi_4\}$, with $\Pi_0 + \Pi_1 + \Pi_2 = \mathbb{1}$ and $\Pi_0 + \Pi_3 + \Pi_4 = \mathbb{1}$. Here, the Born's rule guarantees that $p(1|1) = \mathrm{tr}\{\Pi_0\, \rho\} = p(2|2)$, which renders $p(1|1) = p(2|2)$ for every quantum state. Hence, $(1|1)$ and $(2|2)$ can be thought of as representing the same operational situation.

How to identify the operational equivalences pertinent to each measurement scenario in general theories is still an open question. But this is at the same level of how to identify the joint measurability of observables: it can be done (in principle) in quantum theory, we don't know how to tackle it in general ones, and it can only be certified experimentally in few cases. AFLS hence takes these operational equivalences as a given, as their starting point, and develop the formalism from them.

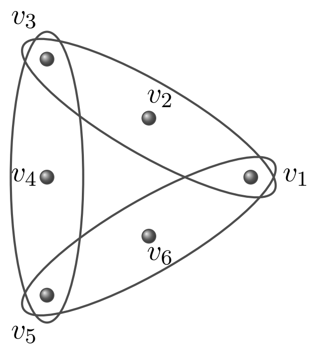
Example 6.1. As a case-study example, consider the situation depicted in Fig. 6.1. This scenario consists of (see Fig. 6.1(a)) three measurements with three outcomes each. Moreover, the operational equivalences that are observed are the following: $v_1 \equiv v_1'$, $v_3 \equiv v_3'$ and $v_5 \equiv v_5'$. Hence, the situation may as well be depicted as in Fig. 6.1(b), where the operationally equivalent events become shared outcomes by the measurements. For the purpose of this lecture, we will refer to the scenario of Fig. 6.1(b) as H_6.

In the previous example I kind of hinted the way that AFLS chooses to represent a contextuality scenario: a hypergraph $H = (V, E)$. The vertices $v \in V$ correspond to the events in the scenario, and the hyperedges $e \in E$ are sets of events representing all the possible outcomes given a particular measurement choice (which is assumed to be complete).

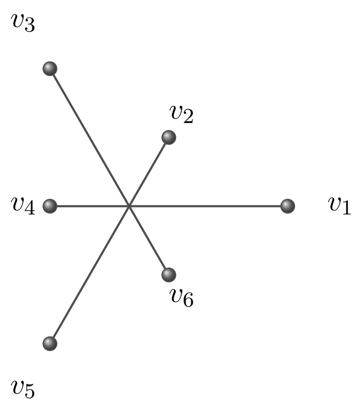
A *probabilistic model* (a.k.a. behaviour) on a contextuality scenario is an assignment of a number to each of the events, $p : V \to [0, 1]$, which denotes the



(a) Scenario with three measurements of three outcomes.



(b) Scenario after the identification via operational equivalences.



(c) The non-orthogonality graph of the scenario.

Figure 6.1: (a) A contextuality scenario with three measurements of three outcomes. The following operational equivalences have been identified: $v_1 \equiv v_1'$, $v_3 \equiv v_3'$ and $v_5 \equiv v_5'$. (b) The representation of the scenario once the operational equivalences are taken into account: measurements 'share' outcomes. (c) The non-orthogonality graph of the scenario.

probability with which that event occurs when a measurement $e \ni v$ is performed. This definition assumes that in the set of experimental protocols we are interested in, the probability for a given outcome is independent of the measurement that is performed. Since the measurements are complete, every probabilistic model p over the contextuality scenario H satisfies the normalisation condition $\sum_{v \in e} p(v) = 1$ for every $e \in E$. The set of all possible probabilistic models on a contextuality scenario is denoted by $\mathcal{G}(H)$.

Example 6.2. Following the case study for scenario H_6, the assignment $p(v_1) = 1 = p(v_4)$ and $p(v) = 0$ for all other vertices is a well defined probabilistic model on H_6, and hence belongs to $\mathcal{G}(H_6)$.

In addition to imposing normalisation, the hyperedges also define the notion of orthogonality (a.k.a. exclusiveness) among events: v and w are orthogonal whenever there exists a hyperedge e that contains both. Hence, in AFLS scenarios, exclusiveness arises as a consequence of the operational equivalences, and does not need to be imposed ad-hoc.

Several sets of probabilistic models have been studied, depending on the nature of the system and the rule used to assign probabilities. The first relevant is that of *classical* models, which comprises the idea of noncontextual deterministic hidden variables. The formal definition is as follows.

Definition 6.3. ***Classical models***
*Let H be a contextuality scenario. An assignment of probabilities $p : V(H) \to [0, 1]$ is a **classical model** if and only if it can be written as*

$$p(v) = \sum_\lambda q_\lambda p_\lambda(v), \tag{6.1}$$

where the weights q_λ and deterministic models p_λ satisfy

$$\sum_\lambda q_\lambda = 1 \qquad \text{and} \qquad p_\lambda(v) = \{0, 1\} \quad \forall\, v, \lambda. \tag{6.2}$$

Since, for finite H, there are only finitely many deterministic models, the set of classical models is a polytope. We denote this polytope by $\mathcal{C}(H)$.

Example 6.4. Following the case study for scenario H_6, the probabilistic model of Example 6.2 is a deterministic one, and hence further belongs to $\mathcal{C}(H_6)$.

AFLS further includes in its formalism the notion of *quantum models*. These are those probabilistic models which can arise in a world complying with the laws of quantum theory. The formal definition is as follows.

Definition 6.5. ***Quantum models***
Let H be a contextuality scenario. An assignment of probabilities $p : V(H) \to [0,1]$ is a quantum model if and only if there exists a Hilbert space $\mathcal{H}$ upon which live some positive-semidefinite quantum state ρ and projection operators P_v associated to every $v \in V$ such that

$$1 = \mathrm{tr}\,(\rho)\,, \quad p(v) = \mathrm{tr}\,(\rho P_v)\ \forall v \in V(H), \ \ \textit{and} \quad \sum_{v\in e} P_v = \mathbb{1}_{\mathcal{H}}\ \forall e \in E(H). \tag{6.3}$$

The set of all quantum models is the ***quantum set****, denoted $\mathcal{Q}(H)$.*

On the one hand, notice that the quantum set is convex. Moreover, it is important to note that the dimension of H is not fixed in the definition of quantum model. In general, H can be infinite-dimensional.

On the other hand, notice that the the definition of quantum models relies on a realisation of the measurements in terms of projectors. That is, general POVM realisations are not considered 'quantum' in the AFLS formalism. This may be seen as a big limitation of the framework, but actually it is a natural restriction: quantum models are asked to be projective just as classical models are demanded to be deterministic (or, in both cases, convex combinations thereof). In a sense, the fact that the framework does not cover classical models defined via indeterministic noncontextual hidden variables, it also must not cover quantum POVM models.

Note further that any contextuality scenario H where $\mathcal{C}(H) = \emptyset$ and $\mathcal{Q}(H) \neq \emptyset$ provides a state-independent proof of the Kochen-Specker theorem. An example of this is given in Fig. 6.2.

AFLS then, in a similar flavour to CSW, uses tools from graph theory to determine whether a particular behaviour is explainable within a noncontextual hidden variable theory. To do so, it further defines the non-orthogonality graph $\mathrm{NO}(H)$ associated to a scenario H as follows: the vertex set of NO is that of H, i.e. V, and two vertices v, u share an a edge in NO if they are not orthogonal in H.

Example 6.6. Following the case study for scenario H_6, its non-orthogonality graph is given in Fig. 6.1(c).

Then, AFLS proves the following: $p \in \mathcal{C}(H)$ iff $\alpha^*(\mathrm{NO}, p) = 1$, where α^* denotes the weighted fractional packing number[1]. Unfortunately, no similar statement can be said for quantum models, i.e. there is no weighted graph invariant of NO that can tell membership to the quantum set. Other interesting sets of behaviours do however relate to graph invariants, and I will present them later on.

[1] The formal definition of α^* goes beyond the scope of these lectures, and I suggest the student reads Appendix A of [69] for an introduction to graph invariants.

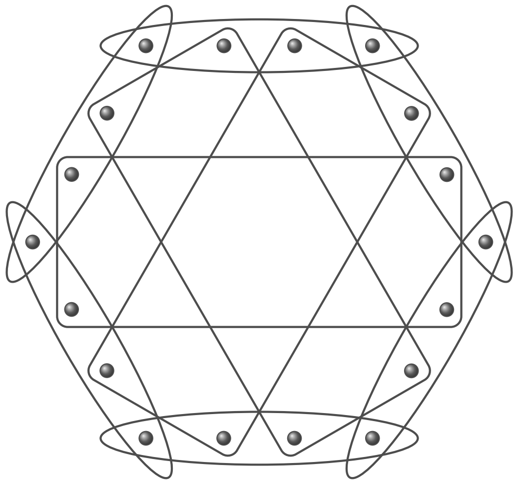


Figure 6.2: The contextuality scenario H_{KS} proving the Kochen-Specker theorem [70, 71]. It has $\mathcal{C}(H_{\mathrm{KS}}) = \emptyset$ but $\mathcal{Q}(H_{\mathrm{KS}}) \neq \emptyset$. The quantum realisation is given by projectors assignments in a Hilbert space of dimension 4.

6.2 Bell scenarios

One of the advantages of the dual approach of AFLS to that of CSW is that it takes into account from the very beginning appropriate normalisation constraints that allows for the direct study of Bell scenarios in the proper manner. Now I will discuss how to actually do it. So basically the question is, given a Bell scenario (n, m, d), how to define a corresponding hypergraph $H = \mathcal{B}_{n,m,d}$ such that $\mathcal{C}(H)$ corresponds to the Bell polytope, $\mathcal{Q}(H)$ to the set of quantum correlations, and $\mathcal{G}(H)$ to the No Signalling polytope.

For the sake of this course, let me present the case study for the CHSH scenario, i.e. $(n, m, d) = (2, 2, 2)$. The subtleties on multipartite Bell scenarios may be found in Sec. 3.3 of [69].

The situation is then as depicted in Fig. 6.3. There, a bipartite Bell experiment is seen as a single-system one, where the inputs are labelled by the pair xy and the joint outcome by the collection ab. The set of events is then defined by $V = \{(ab|xy)\}_{a,b,x,y}$. The tricky bit is to see how to define the hyperedges in $\mathcal{B}_{2,2,2}$. A natural choice would be $E = \{e_{xy} : x, y = 0, 1\}$, where $e_{xy} = \{(ab|xy) : a, b = 0, 1\}$. If we do so, then the following assignment of probabilities would be a valid probabilistic model:

$$p^*(ab|xy) = \begin{cases} 1 & \text{if} \quad (ab|xy) = (00|00)\,, (10|01)\,, (00|10)\,, (00|11)\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Indeed, this assignment is well normalised for each hyperedge e_{xy}. However, when we compute the marginal distribution for Alice's measurements we find the following:

$$p_A^*(0|0) = \sum_b p^*(0b|00) = 1 \neq 0 = \sum_b p^*(0b|01) = p_A^*(0|0)\,.$$

That is, the single party marginals are not well defined, i.e. p^* is a signalling correlation.

The fact that the set of general behaviours for this hypergraph includes Bell signalling ones, however, is not a limitation at the level of the formalism. Rather it is a signal that we were not cautious enough when defining the set of hyperedges in a Bell hypergraph. One could also consider, for instance, what AFLS calls *correlated measurements*, which are also known as *wirings*. Indeed, in Bell scenarios it is the No signalling principle the one that allows for wiring operations to be well defined. Hence, it is not surprising that correlated measurements play a role in defining $\mathcal{B}_{2,2,2}$.

A (bipartite) correlated measurement then is defined by

- a temporal order of the parties. E.g. $A \to B$.
- a choice of measurement for the first party, e.g. x.
- a function $y = f(a)$ for the second party, that determines its measurement input as a function of the previous party's outcome.

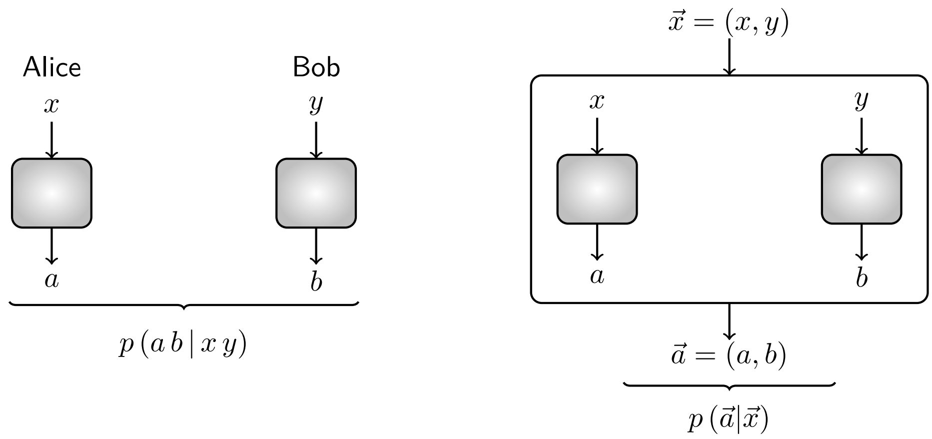
For instance, the choices ($A \to B$, $x = 0$, $y = a$) define a correlated measurement whose four possible outcomes are $\{(00|00), (01|00), (10|01), (10|01)\}$. We already see that a hypergraph containing those four events as a hyperedge would render p^* as not allowed in the set of general behaviours.

Hence, the proposal to define the full set of hyperedges for $\mathcal{B}_{2,2,2}$ is the following:

$$E = \cup_{xy}\{e_{xy}\} \cup_{A\to B} \{e_{\to}\} \cup_{A\leftarrow B} \{e_{\leftarrow}\}\,.$$

With this, it can be proven that $\mathcal{G}(\mathcal{B}_{2,2,2})$ is the CHSH nonsignalling polytope. The hypergraph $\mathcal{B}_{2,2,2}$ is depicted in Fig. 6.4.

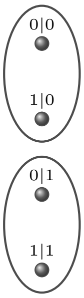
Somehow, it might seem like a paradox that signalling measurements are the ones that ultimately rule out signalling correlations. But actually, it should be thought like this: the fact that a correlated measurement $A \to B$ can be defined is only possible in a scenario where there is no signalling from Bob to Alice. Hence



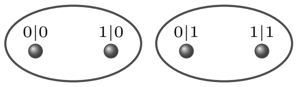
(a) A bipartite Bell experiment

(b) A single system experiment

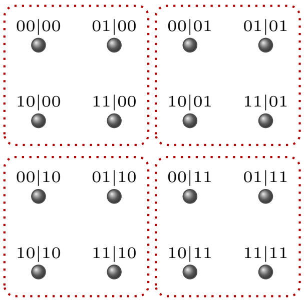
Figure 6.3: (a) Bell experiment: Alice inputs her measurement choice x on her measurement apparatus, depicted by a black box, and obtains an outcome a. Bob similarly inputs y and obtains b. By performing this many times on identical and independent copies of the shared system, Alice and Bob can compute the conditional probability distribution $p(ab|xy)$. (b) A Bell experiment thought of as a single system experiment.



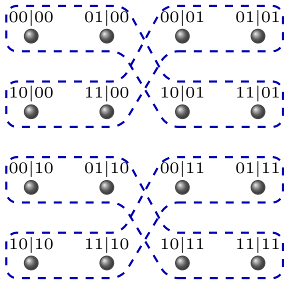
(a) Alice's two binary measurements $B_{1,2,2}$.



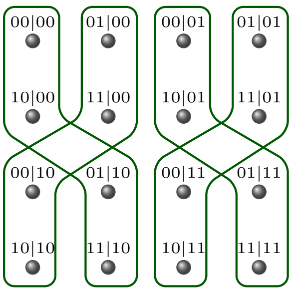
(b) Bob's two binary measurements $B_{1,2,2}$.



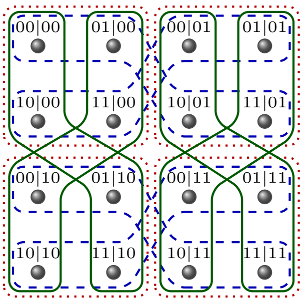
(c) Simultaneous measurements.



(d) Bob's measurement choice depends on Alice's outcome.



(e) Alice's measurement choice depends on Bob's outcome.



(f) Foulis-Randall product: the CHSH scenario $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$.

Figure 6.4: Construction of the CHSH scenario $B_{2,2,2}$ as a Foulis-Randall product $B_{2,2,2} = B_{1,2,2} \otimes B_{1,2,2}$.

imposing that correlated measurements may be plausible indirectly enforces the No Signalling principle.

As a remark, the hypergraph $\mathcal{B}_{2,2,2}$ we've just obtained is actually what in graph theory is called the *Foulis-Randal* product of two hypergraphs, each representing the single-party measurement scenario as depicted in Figs. 6.4(a) and 6.4(b).

As a second remark, consider now the set of quantum models $\mathcal{Q}(\mathcal{B}_{2,2,2})$. Their definition now actually accords with the usual one of quantum correlations[2] [69]. That is, each global measurement represented by the projectors $\{P_v\}_{v\in e}$ can consistently be expressed as a product of local projectors, one for each party, such that the projectors for different parties commute and are properly normalised. In other words, each global projector $P_{ab|xy}$ can be equivalently expressed as $P_{ab|xy} = P_{a|x}P_{b|y}$, where $[P_{a|x}, P_{b|y}] = 0$ for all a, b, x, y and $\sum_a P_{a|x} = \mathbb{1}_{\mathcal{H}}$ (similarly $\sum_b P_{b|y} = \mathbb{1}_{\mathcal{H}}$).

Finally, it can also be shown that $\mathcal{C}(\mathcal{B}_{2,2,2})$ is the traditional Bell polytope. This is clear since one way to define the Bell polytope is as the convex hull of deterministic models, and a deterministic model in the contextuality scenario $\mathcal{B}_{2,2,2}$ is the same as a local deterministic model in the Bell sense. The proof of this claim follows from Props. 4.1.4 and 4.3.1 in [69].

6.3 Compatible observables approach as an events-based scenario

As we are reaching the end of these lectures, let me go back to our starting point: contextuality scenarios that are defined in terms of commuting observables. Now that we have been through these formalisms that heavily rely on graph-theoretic notions, how can we embed the original approach in them? Here I will show you how to do such thing from the point of view of AFLS.

The first thing to notice is that Bell scenarios are actually compatible-observables scenarios: the premise is that Alice's measurements are jointly measurable with those of Bob (guaranteed by their space-like separation), although nothing is assumed between Alice's measurements or Bob's. Hence, we already have a starting point to understand the set of events and 'measurements' of a compatibility scenario within the AFLS approach.

So the main question is, given a contextuality experiment defined in terms of compatible observables, how to construct the events-bases hypergraph from the AFLS formalism. As was mentioned before, a compatible-observables scenario can be depicted as a hypergraph, where the vertices represent the observables, and the

[2] More precisely, with that of quantum correlations in the commutativity paradigm,

hyperedges collect the compatible ones. Recall from Sec. 5.4 that these scenarios are also known as *marginal scenarios*, specified by the triplet $(X, O, \mathcal{M})$. For the sake of simplicity, here I will refer to the marginal scenario solely by X, and to its events-based hypergraph by $H[X]$.

Example 6.7. Let us take as a case study for this course the first formulation of the KCBS scenario. There, the compatibility relations are given in terms of pairs of observables only, hence their compatibility structure is represented by a graph, actually depicted in Fig. 5.2(a). The triplet that defines this marginal scenario is hence: $X = \{A_1, A_2, A_3, A_4, A_5\}\,, \quad O = \{1, -1\}\,, \quad \mathcal{M} = \{\{A_i, A_{i+1}\} : 1 \leq i \leq 5\}$.

So let us start with the vertices. Similarly to what we have intuitively done before, the set of events in $H[X]$ is given by $V = \{(s, C) : C \in \mathcal{M}\,,\ s \in O^C\}$. That is, an event is labelled by a context C and an assignment of an outcome to each observable in C.

Example 6.8. Following the case study, in KCBS each context consists of two observables, hence $O^C = \{(1,1), (1,-1), (-1,1), (-1,-1)\}$. Replacing in the above definition we obtain

$$V = \{(ab, A_i A_{i+1}) : a, b = \pm 1\,,\ 1 \leq i \leq 5\}\,,$$

and hence we recover what we used last lecture.

Now, the convoluted part is that of properly defining the set of hyperedges. In the particular case of Bell scenarios, that included the notion of correlated measurements between parties, i.e. protocols involving compatible observables. In a more general case, where there is no space-like separates structure we cannot rely in such notion to define these extra hyperedges. Instead, the way to do so is to define *measurements protocols* as follows:

Definition 6.9. *Measurement protocol.*
The notion of a measurement protocol is defined recursively, and we refer the reader to the proper mathematical statement D.1.3 of [69]. The idea, based on the notion of measuring compatible observables in a temporal orderly manner, is the following: each measurement protocol is defined first of all by a choice $A \in X$ of observable to use as starting point. Then, by a function f that assigns to each outcome of A an observable that is compatible to it. Then, the measurement protocol continues by defining for each observable $f(a)$ another function f_a that assigns to each outcome of $f(a)$ an observable different from A that is compatible with both A and $f(a)$. And so on. Whenever we reach a point where there are no remaining observables that are compatible with the ones in that branch, the protocol ends.

Somehow, the structure of a measurement protocol resembles that of a decision tree.

Example 6.10. Let us unravel the structure of a generic measurement protocol in the KCBS scenario. Let A_k be our starting observable. Now, a function f should tell us, depending of the outcome of A_k, which is the compatible observable that we are to measure next. So, for this scenario, $f : \{1, -1\} \to \{A_{k-1}, A_{k+1}\}$. There are four such functions, and each will define a different measurement protocol. Note that after the observable $f(a_k)$ is measured, there are no remaining observables that are compatible with both A_k and $f(a_k)$. Hence, the protocol ends here. We see then that the geometry of the scenario renders each branch of each measurement protocol to have two-steps, i.e. involve only two observables.

For the particular choice of $A_k = A_1$ and f s.t. $f(a) = A_{1+a}$, the possible outcomes of this protocol are

$$\{(1, 1|A_1, A_2), (1, -1|A_1, A_2), (-1, 1|A_1, A_5), (-1, -1|A_1, A_5)\}.$$

As a remark, note that when we start from a Bell scenario, these measurement protocols actually describe the correlated measurements I presented before.

We see then that, at least in the the particular case of the KCBS scenario, the non-orthogonality graph of $H[X]$ does coincide with the (complement of) the exclusivity graph of Fig. 5.2(c). Indeed, one can check that for every pair of orthogonal events in the exclusivity graph, there is a measurement protocol that has them both as outcomes. This fact actually generalises to any contextuality scenario $H[X]$.

6.4 Other relevant sets of behaviours

So far we have discussed classical, quantum and general probabilistic models. But other physically relevant sets may as well be defined, and happen to be very well suited to this formalism.

The first one is that of Q_1 *models*. These are of utmost relevance due to their high resemblance to quantum ones. Formally, they are defined as follows:

Definition 6.11. $\mathcal{Q}_1$ ***models***
Let H be a contextuality scenario. An assignment of probabilities $p : V(H) \to [0, 1]$ is a $\mathcal{Q}_1$ model if and only if there exists a Hilbert space $\mathcal{H}$ upon which live some positive-semidefinite quantum state ρ and projection operators P_v associated to every $v \in V$ such that

$$1 = \mathrm{tr}(\rho), \quad p(v) = \mathrm{tr}(\rho P_v) \ \forall v \in V(H), \ \ and \ \ \sum_{v \in e} P_v \leq \mathbb{1}_{\mathcal{H}} \ \forall e \in E(H). \tag{6.4}$$

The set of all these models is denoted $\mathcal{Q}_1(H)$.

Note that the only difference to quantum behaviours arises from the normalisation of the projection operators: here we allow them to be sub-normalised as long as their probabilities sum up to one. Indeed, maximising the value of a noncontextuality inequality over Q_1 models is precisely what CSW do when upper-bounding the value of Bell inequalities violations by quantum models.

Within the AFLS formalism, moreover, a probabilistic model p belongs to the Q_1 set if and only if the weighted Lovász number $\vartheta(\mathrm{NO}, p) = 1$. In contrary to CSW, here the graph whose Lovász number we compute is the non-orthogonality one (i.e. the complement of the exclusivity graph), and the weights are given by the values of the probabilities rather than the coefficients of an inequality.

Computing membership to the Q_1 set, as well as optimising linear functionals on it, are then efficient tasks to accomplish, since they can be cast as SDPs. Whenever the contextuality scenario is that of a Bell scenario, the set of Q_1 models coincide with that of almost quantum correlations [52] we discussed in Sec. 4.2.

Finally, the AFLS formalism is particularly well suited to study sets of behaviours that comply with the Consistent Exclusivity principle (CE). Unfortunately we do not have time in these lectures to discuss the principle in depth, but in a nutshell CE imposes constraints on probabilistic models via conditions that sets of mutually orthogonal events must satisfy. The close connection between the AFLS formalism and graph theory then allows to make the following claims.

Proposition 6.12. *A probabilistic model $p \in \mathcal{G}(H)$*
- *satisfies CE at the single-copy level iff $\alpha(\mathrm{NO}(H), p) = 1$,*
- *satisfies CE at the n-copy level iff $\alpha(\mathrm{NO}(H)^{\boxtimes n}, p^{\otimes n}) = 1$,*
- *satisfies CE at any-copy level iff $\Theta(\mathrm{NO}(H), p) = 1$,*

where $\boxtimes$ denotes the strong product of graphs, α is the weighted independence number and Θ the weighted Shannon capacity.

As a remark, the fact that $\Theta(\mathrm{NO}, p) \leq \vartheta(\mathrm{NO}, p)$ provides a graph-theoretical proof that Q_1 models (and hence almost quantum correlations in Bell scenarios) satisfy the CE principle (comply with the Local Orthogonality principle). In addition, the intuition that a NO graph with the property $\alpha = \Theta < \vartheta$ is related to the existence of a scenario H such that Q_1 is strictly contained in CE^{∞} (i.e. the set of models that comply with CE at any copy level), examples of such situations were found.

I know that this discussion of CE and its implications is very shallow, but I just wanted to give you examples where the tools of graph theory become physically relevant within the AFLS formalism. Those of you who are curious to learn more may find all the details in Section 7 of [69].

6.5 Macroscopic Noncontextuality

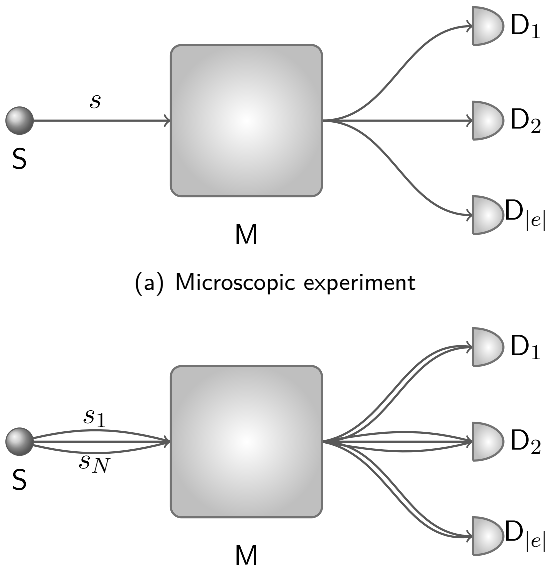
Macroscopic Noncontextuality (MNC) is a principle that was proposed to constrain the set of allowed behaviours in a contextuality scenario [45]. Since a Bell scenario can be viewed as a contextuality one in the AFLS formalism, MNC imposes constraints as well on the correlations allowed in a Bell type experiment. Since the definition of a Bell hypergraph, however, is not done in a 'device independent manner' (i.e. it requires the certification of operational equivalences in the measuring devices), MNC has a slightly different fundamental flavor than the principle that motivated it: Macroscopic Locality. This extra power that comes from the properties of operationally equivalent events is what makes MNC stronger than ML, as we will see bellow.

So let us begin by stating what the principle of MNC is, by following the presentation of [45]. Similarly to ML, MNC demands that a certain "macroscopic limit" of a contextuality experiment has a classical explanation in terms of a deterministic NCHV model. The contextuality scenarios that we have discussed so far constitute the so called 'microscopic' version of the experiment (see figure 6.5): a source S prepares a system s, who enters a measurement device. This device performs a measurement $e \in E$ on s, and sends the particle to one from a set of detectors. The clicking of detector D_v implies that the measurement e yielded outcome $v \in e$. The probability with which detector D_v clicks is then given by $p(v)$, where p is the probabilistic model that characterises the statistics of the experiment.

To define a macroscopic extension of the experiment, consider the following situation. The source now produces N independent copies of this system s, and that these N systems reach the measurement device (see Fig. 6.5). Assume now that we are no longer able to distinguish individual outcomes, but rather the fraction of instances (i.e. "intensity") of each outcome v given a measurement e. The experimental results for a particular measurement in the macroscopic experiment are thus described by a probability distribution $\mathcal{P}(\{I^v\}_{v\in e})$ where I^v denotes the intensity for outcome v. The probabilities for the macroscopic extension are determined by the microscopic probabilistic model $p(v)$, in a way that we will make explicit below.

MNC then imposes that in the limit of large N there exists a non-contextual model for this experiment: the probabilities are such that the intensities for *all* of the outputs v could have been predetermined before the measurement is performed, and the experiment "merely reveals" the intensities that are measured.

So let us denote by I_e^v the random variable associated to variable I^v in the distribution $\mathcal{P}$ when measurement e is performed. A macroscopic experiment is then defined from N "runs" of the microscopic experiment. Let $d_{i\,e}^{v}$ be a random variable that is 1 if v is obtained in the ith run of experiment e and 0 otherwise. The



(a) Microscopic experiment

(b) Macroscopic experiment

Figure 6.5: Pictorial representation of a contextuality experiment from the viewpoint of the Macroscopic noncontextuality principle. (a) Microscopic experiment: A source S prepares a system s, which is sent to the measurement device M. There, an interaction between the measurement apparatus and the system sends the system towards one of a set of detectors, where its presence can be observed as a "detector click". The clicking of detector D_v corresponds to obtaining outcome v. (b) Macroscopic experiment: A source S prepares N independent copies of a system s, which are sent to the measurement device M. There, for each system (and independently for each system), an interaction between the measurement apparatus and the system sends the system towards one of a set of detectors, However, in this case, rather than a single click, there is a distribution of 'clicks' over the detectors according to the probabilities for each outcome in the microscopic experiment. Hence, the 'output' of this macroscopic experiment is the collection of intensities I_e^v registered at the detectors.

intensity of outcome v given measurement e, I_e^v, is then proportional to $\sum_{i=1}^N d_{i\,e}^v$, and its deviation from the mean value is expressed as:

$$\overline{I}_e^v = \sum_{i=1}^N \frac{\overline{d}_{i\,e}^v}{\sqrt{N}} = \sum_{i=1}^N \frac{d_{i\,e}^v - p(v)}{\sqrt{N}}. \tag{6.5}$$

The variables $\overline{I}_e^v$ then are restricted to two natural constraints:

- The sum of the number of hits for all outcomes over all runs must be N, i.e.

$$\sum_{v \in e} \overline{I}_e^v = 0 \quad \forall\, e. \tag{6.6}$$

- In the limit $N \to \infty$, the central limit theorem implies that the probability distribution over the intensity fluctuations for each experiment converges to a multivariate Gaussian distribution.

This last constraint imposes the existence of a covariance matrix γ^e for the experiment e, defined for all $u, v \in e$, with entries satisfying

$$\gamma_{uv}^e = \langle \overline{I}_e^u \overline{I}_e^v \rangle = \langle \overline{d}_{1\,e}^u \overline{d}_{1\,e}^v \rangle = \delta_{uv} p(v) - p(u)p(v). \tag{6.7}$$

Note that the value of γ_{uv}^e is the same for fixed u and v, for any value of e.

Now, MNC imposes that $\{\mathcal{P}(\{I^v\}_{v\in e})\}_e$ (i.e. the collection of the conditional probability distributions $\mathcal{P}(\{I^v\}_{v\in e})$) can be explained in terms of a classical behavior. That is, there should exist a classical distribution over $\{I^v\}$ that recovers $\mathcal{P}(\{I^v\}_{v\in e})$ as marginals, where now the random variables I_e^v cannot depend on the subscript e. In turn, this implies that for this classical behavior over $\{I^v\}$ there must exist a bigger matrix γ_{uv} defined for *all* $u, v \in V(H)$ that acts as its covariance matrix. This γ_{uv} must also have the following properties:

- γ_{uv} must reduce to (6.7) when u, v are restricted to e.
- even for u not in the same measurement as v,

$$\sum_{u \in e} \gamma_{uv} = \langle (\sum_{u \in e} \overline{I}^u) \overline{I}^v \rangle = 0. \tag{6.8}$$

By combining these constraints on γ_{uv} with the relations between the microscopic behavior $p(v)$ and the covariance matrix γ_{uv}^e one can characterize the probabilistic models that are compatible with MNC as follows.

Theorem 6.13. *[45]*
A probabilistic model p on scenario H is macroscopically non-contextual if there exists a "macroscopic non-contextuality certificate": a p.s.d. matrix γ ranging over all $v \in V(H)$ such that

- $\sum_{u \in e} \gamma_{uv} = 0$;
- $(u, v \in e$ *and* $u \neq v) \Rightarrow \gamma_{uv} = -p(u)p(v)$;
- $\gamma_{vv} = p(v) - p(v)^2$;

The key point is that these constraints are equivalent to those that appear in the alternative definition of the set of $\mathcal{Q}_1$ models of Def. 6.11:

Definition 6.14. $\mathcal{Q}_1$ ***models*** *[45]*
Given a scenario H, a probabilistic model p belongs to the set $\mathcal{Q}_1$ iff there exists a moment matrix $M \geq 0$ ranging over all $v \in V(H)$, and a special extra column labelled 1, such that

1. $\sum_{u \in e} M_{uv} = P(v)$ *for all* $u \in V(H)$;
2. $(u, v \in e$ *and* $u \neq v) \Rightarrow M_{uv} = 0$;
3. $M_{vv} = P(v)$;
4. $M_{1v} = P(v)$ *and* $M_{11} = 1$;

The proof of the equivalence is left as an exercise, and may be found in [45].

When Bell scenarios are considered, the set of $\mathcal{Q}_1$ models coincides with that of almost quantum correlations, which is a strict subset of the first level of the NPA hierarchy. Hence, MNC imposes stronger constraints than ML in these scenarios. The main difference between the formulations of MNC for Bell scenarios and ML is that in the former (i) the events in the moment matrix involve global outcomes, i.e. the alphabet used to label the rows and columns of the matrix has only letters that correspond to joint outcomes, and (ii) one-way LOCC measurements (i.e. correlated measurements) are consider as feasible actions in a Bell scenario, hence they effectively impose extra linear constraints on the moment matrix. The fact that in the hypergraph formulation of a Bell scenario correlated measurements are feasible actions is implied by the operational equivalences between events. This is not a property that can be certified in a device independent manner, hence MNC is not ultimately a device-independent principle to constraint correlations.

7 Computation toolbox and many-body nonlocality

La primer parte de esta unidad está enfocada a presentar las herramientas computacionales para estudiar correlaciones. A través de ejemplos, se muestran scripts para calcular puntos extremos de polytopos, desigualdades de Bell/contextualidad, y la caracterización de las correlaciones casi-cuánticas mediante programas semidefinidos (SDP). Luego se presenta el problema de la "intractabilidad de sistemas con escenarios grandes", y se mencionan propuestas (teóricas y experimentales) para detectar nolocalidad en many-body systems.

7.1 Computational tools

Here we will see how to use computational tools to compute facet-defining Bell inequalities and estimate Tsirelson's bound. We will also see how to check for membership to the Classical and Almost Quantum sets.

7.1.1 The Bell polytope

The Bell polytope is defined either by (i) the complete list of its extreme points, or (ii) the complete list of its facets. The extreme points correspond to the deterministic correlations, and the facets are Bell inequalitites. Listing all the deterministic strategies for a given (n, m, d) scenario is indeed an easy task, hence it is the starting point to study the Bell polytope.

Exercise 7.1. Outline a script that lists the deterministic strategies.

5-10 minutes to think about this

A script that lists the extreme points may consist of the following:

- List the d^m single-party strategies.
- Generate all the n-tuples with elements from $\{1, \ldots, d^m\}$
- For each of those n-tuples, create the deterministic p. Store the p as a row in a matrix.

In order to make the computation more efficient, we work with the smallest representation of a non-signalling probability vector, which is given by its Collins-Gisin form (see eq. (3.4)).

Exercise 7.2. Implement such a script in, say, Matlab.

Once we have the list of extreme points, there are (at least) two options to proceed. On the one hand we can directly tackle the problem of whether a probability vector has a classical realisation. On the other hand, we can also use these extreme points to find the equations of the facets that define their convex hull.

For the former, all we need to check given a conditional probability distribution is whether it can be decomposed as a convex combination of those extreme points.

Exercise 7.3. Outline a script that checks if a probability vector has a classical model.

5-10 minutes to think about this

A script that performs such a task may be as follows:

- list the extreme points D_p
- declare the variable of coefficients: c
- list the constraints on c: normalization, positivity, and reproduces the statistics via $p == c\,D_p$.

This is a feasibility problem rather than an optimisation one. There are different optimisation toolboxes for these types of optimisations, and here we will use `CVX` [72, 73] with either the `sdpt3` [74, 75] or the `sedumi` [76] solvers.

Exercise 7.4. Implement such a script in Matlab for an arbitrary Bell scenario. Check how the elapsed-time for the computation changes with the Bell scenario. Test: deterministic points or totally uncorrelated probability vector.

We can also use the list of extreme points to find the equations of the facets that define their convex hull. There are several software packages that do this, such as `PORTA` [77], `cdd` [78] and `PANDA` [79].

Exercise 7.5. Take a CHSH Bell scenario. Compute the list of extreme points with the script. Then, input the list of extreme points in one of these packages and find the facets. Take other scenarios, check if it's feasible to find the facets and when it starts being intractable.

7.1.2 Amost quantum correlations as an SDP

As was mentioned in the previous lectures, how to test whether a probability vector has a quantum realisation is not yet known. The best we can do up to date is to test the NPA hierarchy. The almost quantum set is somehow related to this hierarchy as was mentioned before, and is moreover of fundamental physical relevance. Here we will put our hands on coding a simple membership script to this set.

Exercise 7.6. Consider a bipartite Bell scenario. From the definition 4.2 of the almost quantum set as an SDP, outline a script that checks membership.

10-20 minutes to think about this

Similarly to the case of Bell polytopes, `CVX` with `sedumi` and `sdpt3` come in handy when implementing this script.

Exercise 7.7. Implement in Matlab the script that test membership to the almost quantum set for a Bell scenario $(2, m, d)$. Test: deterministic boxes, singlet correlations, PR-boxes.

Now that we know how to tell whether a probability vector is almost quantum or not, we can optimise violations of Bell inequalities by correlations in this set.

Exercise 7.8. Implement a script in Matlab that optimises a linear functional over the set of almost quantum correlations. Test the script for the facet-defining Bell inequalities obtained earlier. Recover Tsirelson's bound for CHSH, and check how the optimisation run-time increases with the size of the scenario.

7.1.3 Adjusting bounds to account for the detection loophole

Let us consider, for the clarity of the presentation, a Bell scenario consisting of two parties (Alice and Bob), i.e. a $(2, m, d)$ Bell scenario. In any real experimental demonstration of nonlocality there will necessarily be experimental imperfections that mean that the idealised treatment of Sections 1 to ch:L3 will not be strictly realised. In particular, in many experimental demonstrations there will necessarily be loss: not every particle pair distributed between Alice and Bob will arrive at their laboratories, and even if they do, their detections will not necessarily always register an outcome. The detection loophole refers to the fact that if one makes the fair sampling assumption for Alice and Bob, i.e if it is assumed that the conclusive events (where no particle is lost) constitute a faithful representative of the complete experimental data, and then apply the idealised treatment to it, then one may erroneously conclude that nonlocality has been demonstrated, even though the underlying state was compatible with a classical description.

So the question is how to account for the effect of losses to properly certify nonlocality? To formally do this, consider a Bell scenario where now every measurement has instead $d+1$ outcomes, i.e. a $(2, m, d+1)$ scenario. The additional outcome, which we here denote by $a = 0$ (resp. $b = 0$), represents the situation where, due to the above mentioned experimental imperfections, no event was registered at Alice's (resp. Bob's) measuring device. Branciard [80] called this the *a priori* scenario, and we will use his notation here. The probability vector $p_0(ab|xy)$ denotes a correlation in the scenario $(2, m, d+1)$ and ultimately represents the whole statistics of the experiment, that is, it takes into the account the probability of 'no click' in each wing of the setup. If we denote by η_x and η_y the experimental efficiencies of the mesurements x and y respectively, these are related to the probability vector components via

$$\begin{aligned}
\eta_{xy} &= \sum_{\substack{a\neq 0\\ b\neq 0}} p_0(ab|xy)\,,\\
\eta_x &= \sum_{a\neq 0} p_0^A(a|x)\,,\\
\eta_y &= \sum_{b\neq 0} p_0^B(b|y)\,,
\end{aligned}$$

where $p_0^A(a|x) = \sum_{b=0}^{d} p_0(ab|xy)$ and $p_0^B(b|y) = \sum_{a=0}^{d} p_0(ab|xy)$ are Alice's and Bob's marginal statistics. Here on for simplicity we collect the efficiencies $\{\eta_{xy}\}_{x,y}$ in the matrix $\boldsymbol{\eta}^{AB}$, $\{\eta_x\}_x$ in the matrix $\boldsymbol{\eta}^{A}$, and $\{\eta_y\}_y$ in the vector $\boldsymbol{\eta}^{B}$. We use the notation $\boldsymbol{\eta} = \{\boldsymbol{\eta}^{AB}, \boldsymbol{\eta}^{A}, \boldsymbol{\eta}^{B}\}$ to refer to the full data on the detection efficiencies, which we assume Alice and Bob estimate in the Bell test.

Now, when one performs a Bell experiments, the recorded data ultimately generates the correlations p_0 in $(2, m, d+1)$, and from them the data from successful runs of the experiments is post-selected and used to compute the correlations p in $(2, m, d)$. The key point is to notice that p_0 and p are related as follows:

$$p(ab|xy) = \frac{1}{\eta_{xy}} p_0(ab|xy)\,.$$

The detection loophole then states the fact that there exist $p_0 \in \mathcal{C}_{n,m,d+1}$ that post-select into a $p \notin \mathcal{C}_{n,m,d}$. What we will see now is how to certify via Bell inequalitites the nonlocality (if any) demonstrated by the data p and $\boldsymbol{\eta}$.

The set of behaviours that arises from the *a priori* local set by postselecting on successful rounds of the experiment is hence given by

$$
\begin{aligned}
\Sigma_{\mathrm{ps}}^{\mathrm{LHV}}(\boldsymbol{\eta}) = \Big\{\{p(ab|xy)\}\Big| \\
p(ab|xy) &= \frac{1}{\eta_{xy}} \sum_{\lambda} D_{\lambda}^{0}(ab|xy) q_0(\lambda) \ \ \forall a,b,x,y, \\
q_0(\lambda) &\geq 0 \ \ \forall \lambda, \ \ \sum_{\lambda} q_0(\lambda) = 1, \\
\eta_{xy} &= \sum_{\substack{\lambda, a\neq 0, \\ b\neq 0}} D_{\lambda}^{0}(ab|xy) q_0(\lambda) \ \ \forall x,y, \\
\eta_{x} &= \sum_{\substack{\lambda, a\neq 0 \\ b}} D_{\lambda}^{0}(ab|xy) q_0(\lambda) \ \ \forall x, \\
\eta_{y} &= \sum_{\substack{\lambda, a \\ b\neq 0}} D_{\lambda}^{0}(ab|xy) q_0(\lambda) \ \ \forall y \Big\}, \qquad (7.1)
\end{aligned}
$$

where $D_{\lambda}^{0}(ab|xy)$ represents a deterministic probability vector in the $(2, m, d+1)$ scenario.

Now, consider a linear Bell functional β specified by the coefficients $\{I_{abxy}\}_{a,b,x,y}$,

$$\beta = \sum_{a,b,x,y} I_{abxy} p(ab|xy). \qquad (7.2)$$

In an ideal scenario (with no losses) the violation of a Bell inequality is the observation of $\beta > \beta_C$, where

$$\beta_C = \max_{\{p(ab|xy)\}\in \mathcal{C}_{2,m,d}} \sum_{a,b,x,y} I_{abxy} p(ab|xy)\,. \qquad (7.3)$$

In the post-selected scenario, with efficiencies $\boldsymbol{\eta}$, the post-selected LHV bound of the functional can also be defined, and is given by

$$\beta_C(\boldsymbol{\eta}) = \max_{\{p(ab|xy)\}\in \Sigma_{\mathrm{ps}}^{\mathrm{LHV}}(\boldsymbol{\eta})} \sum_{a,b,x,y} I_{abxy} p(ab|xy). \qquad (7.4)$$

A value $\beta > \beta_C(\boldsymbol{\eta})$ then provides a detection-loophole-free certification of nonlocality.

Exercise 7.9. Outline a script that checks if a probability vector p belongs to the post-selected set $\Sigma_{\mathrm{ps}}^{\mathrm{LHV}}(\boldsymbol{\eta})$ for a given $\boldsymbol{\eta}$.

5-10 minutes to think about this

Exercise 7.10. Consider a CHSH scenario where in addition the the probability of no-click events in Alice's and Bob's labs are uncorrelated. Consider the symmetric situation where the marginal efficiencies are $\eta_x = \eta\ \forall x$ and $\eta_y = \eta\ \forall y$, hence $\eta_{xy} = \eta^2\ \forall x, y$. Compute $\beta_C(\eta)$ for the CHSH inequality, and recover the known result of Eberhard that a detection-loophole-free test is not possible in this scenario whenever $\eta \leq \frac{2}{3}$ [81].

7.2 Many-body nonlocality

Detecting the nonlocal character of correlations observed in an experiment is an interesting problem. In principle, one needs to consider the local polytope of the corresponding Bell scenario and check whether the conditional probability distribution lies inside or outside of it. However, from a practical point of view this approach is inconvenient for large scenarios, since the dimensionality of the polytope increases exponentially with the number of parties, which makes the problem computationally intractable (as we noticed earlier in the lecture).

In order to tackle the problem of witnessing nonlocality in many-body systems, one hence needs to simplify somehow the problem. On proposal was to focus the study on Bell inequalities that contain only one and two-body correlators [82, 83]. In principle one could argue the relevance of such inequalities, since in general the correlators that involve a large number of parties are those which carry detailed information about the correlations. Contrary to this intuition, one and two-body correlators prove already useful for detecting nonlocality in physically relevant systems [82, 83]. Indeed, one may further restrict the two-body correlators Bell inequalities to those that satisfy certain symmetries regarding the labelling of the parties. In this lecture I will first review the approach of [82, 83], focusing on the permutational invariance symmetry [82].

So far I have made emphasis on discussing correlations in terms of the probability vector, i.e. the the conditional probability distribution $p(ab|xy)$. But as I mentioned in Lecture 1, Bell inequalities can also be written in terms of correlators E_{xy} (see e.q. (1.3)), which is an equivalent representation in the case of dichotomic measurements. Correlators are particularly well suited for the study of this simplified Bell test, hence let us begin by defining some general notation.

Within this representation, the objects that correspond to the probabilistic models on scenario $(n, m, 2)$ are now vectors $\mathbf{M}$, whose components are given by the correlators $\{E_{x_{i_1}\ldots x_{i_k}} : \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\},\, k \leq n\}$. These, may moreover be thought of as *expectation values of physical observables*, just like in section 1.6. Indeed, given a set of dichotomic observables $\{\mathcal{M}_k^{(i)}\}_k$ for each party i, each component of the vector $\mathbf{M}$ can be written as $E_{x_1\ldots x_k} = \langle \mathcal{M}_{x_1}^{(1)} \cdots \mathcal{M}_{x_k}^{(k)} \rangle$.

The components of the vectors $\mathbf{M} \in \mathbb{R}^{(m+1)^n-1}$ are expressed as follows:

- First, $n{*}m$ components which correspond to the $n{*}m$ single-party correlators $\langle \mathcal{M}_k^{(i)} \rangle = P_i(0|k) - P_i(1|k)$. Here, $P_i(a|k)$ denotes the probability that party i obtains outcome a when measuring k. $\mathcal{M}_k^{(i)}$ is usually referred to as the "observable" k measured by party i.
- Second, $\binom{n}{2} m^2$ components which correspond to the two-party correlators $\langle \mathcal{M}_{k_1}^{(i_1)} \mathcal{M}_{k_2}^{(i_2)} \rangle = \sum_{a_1,a_2} (-1)^{a_1 \oplus a_2} P_{i_1 i_2}(a_1 a_2 | k_1 k_2)$.
- Continue, for each $j = 3 \dots n$, with $\binom{n}{j} m^j$ components which correspond to the j-party correlators

$$\langle \mathcal{M}_{k_1}^{(i_1)} \dots \mathcal{M}_{k_j}^{(i_j)} \rangle = \sum_{a_1 \dots a_j} (-1)^{a_1 \oplus \dots \oplus a_j} P_{i_1 \dots i_j}(a_1 \dots a_j | k_1 \dots k_j). \quad (7.5)$$

It is straightforward to check that $\sum_{j=1}^{n} \binom{n}{j} m^j = (m+1)^n - 1$, as the dimension of the vector $\mathbf{M}$. Similar to the case of representing the correlations via the vector of probabilities, for classical models the correlators $\langle \mathcal{M}_{k_1}^{(i_1)} \dots \mathcal{M}_{k_j}^{(i_j)} \rangle$ take a product form $\langle \mathcal{M}_{k_1}^{(i_1)} \rangle \cdot \ldots \cdot \langle \mathcal{M}_{k_j}^{(i_j)} \rangle$. Hence, the set of classical correlations is characterized by the convex hull of the deterministic correlators $\mathbf{M}_D$, defined as those $\langle \mathcal{M}_{k_1}^{(i_1)} \rangle \cdot \ldots \cdot \langle \mathcal{M}_{k_j}^{(i_j)} \rangle$ with local mean values being $\langle \mathcal{M}_{k_l}^{(i_l)} \rangle = \pm 1$. The set of classical correlations is again a polytope that we denote by $\mathbb{P}$. As mentioned in section **??**, the facets of this polytope correspond to the *tight* Bell inequalities of the scenario $(n, m, 2)$.

Most of the known constructions of multipartite Bell inequalities contain highest-order correlators, i.e., those with $j = n$ in eq. (7.5). However, throughout this chapter we will see how to engineer Bell inequalities that witness nonlocality only from one and two body[1] expectation values. In addition, we will focus on the case of two measurements per party. The general form of such a Bell inequality is

$$\begin{aligned} &\sum_{i=1}^{n} (\alpha_i \langle \mathcal{M}_0^{(i)} \rangle + \beta_i \langle \mathcal{M}_1^{(i)} \rangle) + \sum_{i<j}^{n} \gamma_{ij} \langle \mathcal{M}_0^{(i)} \mathcal{M}_0^{(j)} \rangle + \\ &+ \sum_{i \neq j}^{n} \delta_{ij} \langle \mathcal{M}_0^{(i)} \mathcal{M}_1^{(j)} \rangle + \sum_{i<j}^{n} \varepsilon_{ij} \langle \mathcal{M}_1^{(i)} \mathcal{M}_1^{(j)} \rangle + \beta_C \geq 0, \end{aligned} \quad (7.6)$$

where $\alpha_i, \beta_j, \gamma_{ij}, \delta_{ij}$, and ε_{ij} are some real parameters, while β_C is the so-called classical bound.

[1] I will use the words n-party and n-body interchangeably.

7.2.1 Two-body correlators

Now we would like to simplify the problem by considering only one and two-body correlations. That is, from the polytope $\mathbb{P}$ in the complete scenario, we construct a polytope $\mathbb{P}_2$ of classical correlations by neglecting correlators of order higher than two. Indeed, we take all elements $\mathbf{M}$ of $\mathbb{P}$ and simply remove the components which correspond to j-party correlators with $j \geq 3$. Similarly, the vertices of $\mathbb{P}_2$ are those collections of correlators $\mathbf{M}_2$ for which $\langle \mathcal{M}_k^{(i)} \mathcal{M}_l^{(j)} \rangle = \langle \mathcal{M}_k^{(i)} \rangle \cdot \langle \mathcal{M}_l^{(j)} \rangle$, where local mean values are ± 1.

Although $\dim \mathbb{P}_2 = 2n^2$ is much smaller than $3^n - 1$ (the dimension of $\mathbb{P}$), it still grows with the number of parties, which makes difficult the task of determining the facets of $\mathbb{P}_2$. One way to overcome this problem is to restrict the study to Bell inequalities that obey some symmetries. In the following, I will discuss Bell inequalities which are symmetric under permutation of the parties and further contain only one and two-body correlators.

Given a Bell inequality, imposing permutational symmetry means that when we exchange the label (order) of any party the equation remains the same. Mathematically, for a two-body correlators Bell inequality in the form (7.6) this implies that the expectation values $\langle \mathcal{M}_k^{(i)} \rangle$ and $\langle \mathcal{M}_k^{(i)} \mathcal{M}_l^{(j)} \rangle$, with fixed k, l and different i, j, appear in the Bell inequality (7.6) with the same "weights", i.e. $\alpha_i = \alpha$, $\beta_i = \beta$, and so on. Hence, the general form of a symmetric Bell inequality with one- and two-body correlators is

$$I := \alpha \mathcal{S}_0 + \beta \mathcal{S}_1 + \frac{\gamma}{2} \mathcal{S}_{00} + \delta \mathcal{S}_{01} + \frac{\varepsilon}{2} \mathcal{S}_{11} \geq -\beta_C, \tag{7.7}$$

where $\alpha, \beta, \gamma, \delta, \varepsilon$ are real parameters, and $\mathcal{S}_k$ and $\mathcal{S}_{kl}$ (with $k, l = 0, 1$) denote the one- and two-body correlators symmetrized over all observers, i.e.,

$$\mathcal{S}_k = \sum_{i=1}^{n} \langle \mathcal{M}_k^{(i)} \rangle, \qquad \mathcal{S}_{kl} = \sum_{i \neq j = 1}^{n} \langle \mathcal{M}_k^{(i)} \mathcal{M}_l^{(j)} \rangle. \tag{7.8}$$

Geometrically, the polytope $\mathbb{P}_2$ is mapped under permutational symmetry onto a simpler one $\mathbb{P}_2^S$, which independently of the number of parties, is always five-dimensional and it elements are the vectors $(\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_{00}, \mathcal{S}_{01}, \mathcal{S}_{11})$. Note that the number of vertices is significantly reduced, from 2^{2n} for $\mathbb{P}_2$ to $2(n^2+1)$ for $\mathbb{P}_2^S$.

Now, if one follows the line of thought of the previous lectures, it is natural to search for facet-defining Bell inequalities of the form (7.7), by computing the facets of $\mathbb{P}_2^S$. For this, a characterisation of the extremal points of $\mathbb{P}_2^S$ is required. Such a charcterisation is not trivial like in the case of a usual Bell scenario, and goes beyond the scope of this lecture. Those who are interested in the details of this approach may consult [82].

So let us discuss the particular case of a class of permutationally invariant two-body correlators Bell inequality. Its form is given by [82]

$$\begin{aligned} I := x[\sigma\mu \pm (x+y)]\mathcal{S}_0 + \mu y \mathcal{S}_1 + \frac{x^2}{2}\mathcal{S}_{00} + \sigma x y \mathcal{S}_{01} + \frac{y^2}{2}\mathcal{S}_{11} \\ \geq \frac{1}{2} - \frac{1}{2}\left[n(x+y)^2 + (\sigma\mu \pm x)^2\right], \end{aligned} \tag{7.9}$$

where x and y are positive natural number, $\sigma = \pm 1$, and μ is an integer with opposite parity to y^2 for odd n and to x^2 for even n. Although in general these inequalities are usually not facet-defining, they are still useful for detecting quantum nonlocality, as we will see in the following example.

Now set the parameters to be $x = y = -\sigma = 1$, and $\alpha_- = -2$. From eq. (7.9) the classical bound is $\beta_C = 2n$, and the resulting Bell inequality reads

$$I_n = -2\mathcal{S}_0 + \frac{1}{2}\mathcal{S}_{00} - \mathcal{S}_{01} + \frac{1}{2}\mathcal{S}_{11} + 2n \geq 0. \tag{7.10}$$

Quantum violations of ineq. (7.10) can be found by setting all parties to measure the same pairs of observables, i.e., $\mathcal{M}_j^{(i)} = \mathcal{M}_j$ for every $i = 1, \ldots, n$. Without loss of generality, this are parametrised as as $\mathcal{M}_0 = \sigma_z$ and $\mathcal{M}_1(\theta) = \cos\theta\sigma_z + \sin\theta\sigma_x$ for $\theta \in [0, \pi]$. Denote by $I_n(\theta)$ the value of the linear functional as a function of θ, where the state used to compute the correlators is optimised to give the maximum violation. Inequality (7.10) is then violated if there exists θ such that $I_n(\theta) \ngeq 0$. Fig. 7.1 presents the value of $I_n(\theta) \ngeq 0$ for various values of n. Numerically, we see that the effective violation (i.e. the violation divided by the classical bound) grows with n, and becomes more robust against misalignments of θ for large n.

Other approaches

- JD experiment: witnesses
- Jordi's paper

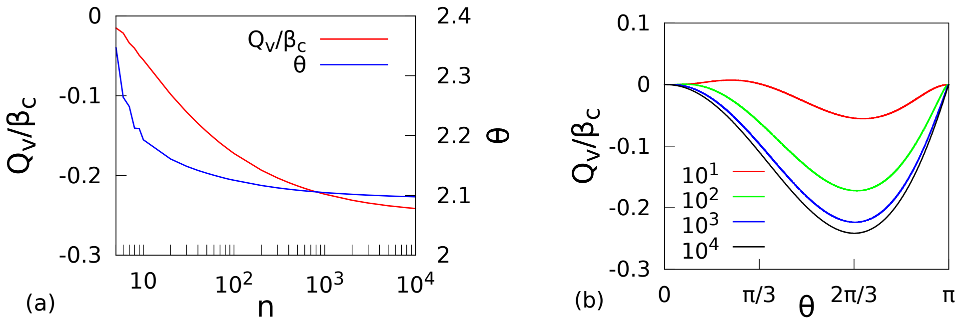


Figure 7.1: (a) The effective (divided by the classical bound) maximal violation of Ineq. (7.10) (red line) and the corresponding angle θ in $\mathcal{M}_1$ (blue line) as functions of n. (b) Effective violation of Ineq. (7.10) as a function of θ for $n = 10^k$ with $k = 1, 2, 3, 4$. For large n the violation is robust against misalignments of the second observable.

8 Applications and open problems

En esta unidad se presentan algunas aplicaciones de los fenómenos de nolocalidad y contextualidad. Primero se discute la necesidad de protocolos "device-independent" en tareas criptográficas, y luego se presentan la distribución cuántica de claves device-independent, la amplificación y expansión de aleatoriedad de manera device-indepentent, y la aplicación de argumentos device-independent para testear la dimensión de espacios de Hilbert.

8.1 Device-independent quantum information: Bell nonlocality as a resource

Bell nonlocality is of particular relevance when studying quantum information protocols in device-independent paradigms. Consider as an example the case of quantum key distribution (QKD). There, Alice and Bob are able to distill a secure key by making measurements on a shared system in an entangled state. The security of the key then relies on quantum properties, such as the "no-cloning theorem" and "information gain implies disturbance" [7]. Now, when analyzing security in this protocols one can be *the most paranoid* and even allow for the case where Alice and Bob can acquire the source and measurement devices from the eavesdropper. The parties involved in the QKD protocol therefore cannot make any physical assumptions on the inner working of their quantum devices. Hence, in this extreme scenario, the only information that Alice and Bob can afford to rely on are their measurement choices (provided they control them) and the classical label of the measurement outcomes. This extreme scenario is what is called the *device-independent paradigm*. QKD protocols formulated in a device-independent way then rely rather on Bell inequality violations to certify security, as we will see next.

Even though the need of device-independent scenarios may be motivated from a cryptography perspective, it is not ultimately limited to it. In this lecture I will discuss a couple of examples.

8.1.1 Device-independent quantum key distribution

In a cryptographic scenario, two honest parties (Alice and Bob) want to exchange information in a private way. However, there is a third party, who is a dishonest eavesdropper (Eve) that wants to read this information. In a cryptographic protocol then, Alice and Bob's task (called 'key distribution') is to establish a secret key, namely two lists (one for Alice and one for Bob) of perfectly correlates bits of which Eve has no information. This secret key is later on consumed in order to encode (and decode) the message in a secure way.

Here we want to study key distribution protocols within the device-independent formalism and with quantum resources. The first of such protocols was introduced by Ekert in his celebrated paper [84]. But in order to understand the relevance of the device-independent paradigm, let us first discuss quantum key distribution, in particular a version of the BB84 protocol [85] for standard QKD. In this protocol, Alice and Bob share an entangled pair of qubits on the maximally entangled state $|\Psi\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$. Using this state, the parties proceed as follows:

- Alice chooses to measure her qubit either in the z basis, i.e. $\{|0\rangle, |1\rangle\}$, or in the x basis, i.e. $\{|+\rangle, |-\rangle\}$ with $|\pm\rangle = \frac{|0\rangle\pm|1\rangle}{\sqrt{2}}$. Whenever she measures in z and obtains 0 or she measures in x and obtains $+$, she maps the measurement outcome to the bit 0. Otherwise, she maps the outcome to the bit 1.
- Similarly, Bob measures his qubit randomly in one of these two bases. He maps his results into a classical bit using the same convention as Alice.
- Alice and Bob publicly announce their choice of measurement bases. If these coincide, Alice and Bob keep their generated classical bits, which are known only to them. This process is known as *basis reconciliation*.

Notice that if Alice's and Bob's choice of basis agree, their generated bits are perfectly correlated, and otherwise completely uncorrelated. Indeed, if Alice measures in the z basis, then Bob's qubit is steered into the state $|0\rangle$ when she obtains outcome 0, and into $|1\rangle$ when she obtains 1. Hence, should he measure in the z basis as well he would obtain the same outcome as Alice. Since both parties use the same convention to map their outcomes into a classical bit, their classical bits will be perfectly correlated. The same situation happens when both parties measure in the x basis. On the other hand, consider the case where Alice and Bob choose different bases, say z and x respectively. Then, since $|\langle \pm z, \pm x\rangle|^2 = \frac{1}{2}$, where $|+z\rangle = |0\rangle$ and $|-z\rangle = |1\rangle$, then their outcomes will be completely uncorrelated, and so too their classical bits. Indeed, this situation is equivalent to Bob discarding his system and flipping a coin. Therefore, after many rounds of the protocol and basis reconciliation, Alice and Bob discard all the bad instances and end up sharing a list of perfectly correlated random bits.

Now, given the no-cloning theorem and monogamy of entanglement, Eve cannot have a qubit prepared on the same state as Bob's. She can even be in possession of the source of entangled qubits sent to Alice and Bob. However, any attempt to tamper with the state of the system will introduce errors in the correlated outcomes between Alice and Bob. By publicly disclosing part of the generated key and noticing the discrepancies, Alice and Bob can then check whether Eve tried to learn their private key. Depending on the number of errors, Alice and Bob can then decide whether to error-correct the remaining private part of their key or to abort the protocol.

This analysis of a particular type of attack by Eve is far from being a complete proof of the general security of Quantum Cryptography, but summarises the main features common to most of the known schemes.

Now, to illustrate the need for device-independent protocols, consider the discussion in [86] about this BB84 protocol. In the noisy free case, the correlations that are observed between Alice and Bob are $p(a = b|A = B) = 1$ and $p(a = b|A \neq B) = \frac{1}{2}$, where a (b) denotes the classical bit to which Alice (Bob) mapped her (his) measurement outcome, and A (B) denotes her (his) choice of measurement basis. If such correlation results from measurements in the z and x bases on qubit pairs, then the state of these two qubits is necessarily maximally entangled (due to self testing) and security follows. However, the same correlation can also be reproduced by the following system and measurement devices: as system, take four qubits on the state:

$$\rho = \tfrac{1}{4}(|00\rangle\langle 00| + |11\rangle\langle 11|) \otimes (|{+}{+}\rangle\langle{+}{+}| + |{-}{-}\rangle\langle{-}{-}|)\,,$$

where Alice holds the first and third qubit, and Bob the other two. The measurement devices are engineered as follows: when Alice inputs her choice of measurement, if this choice is the z (x) basis, the device measures the first (third) qubit in this basis. The same happens for Bob, with the second and fourth qubit. Clearly, their measurement results are completely correlated when the bases agree and uncorrelated otherwise, precisely as for the ideal BB84 case. However, their state is separable, so a secure key cannot be established. The BB84 protocol becomes insecure even in the ideal noise-free situation.

If then we cannot afford to rely in the inner working of the sources and measurement devices, as is the case is an extremely paranoid cryptography scenario, one needs to move to the device-independent paradigm. A typical Device-independent quantum key distribution (DIQKD) protocol consists of the following steps:

- a measurement step, where Alice and Bob measure a series of entangled quantum systems. Equivalently, Alice and Bob use their devices a series of rounds, producing each time an outcome from their devices by selecting the input.

- an estimation step, in which Alice and Bob publicly announce a fraction of their measurement results. With this, they can estimate the violation of a Bell inequality and hence check whether their data could have been produced by a classical device potentially known by Eve. In addition, with this information they can also estimate the error rate in their raw data.
- an error correction step, in which these errors are corrected using a classical protocol that involves public communication.
- a privacy-amplification step in which a shorter, secure key is distilled from the raw key based on a bound on Eve's information deduced from the Bell violation estimation.

It goes beyond the scope of the course to discuss these QKD protocols fully, and I will instead review the protocol by Barrett *et al.* as an example [87]. This protocol is of particular importance among DI ones, since it is the first one to prove security without assuming the validity of quantum theory. That is, they allow for eavesdroppers who can break the laws of quantum mechanics, as long as nothing they can do implies the possibility of superluminal signaling.

So let us begin with the protocol, which generates a single shared secret bit, guaranteed secure against general attacks by postquantum eavesdroppers. For $N \in \mathbb{N}$, define the bases $X_r = \left\{\cos\left(\frac{r\pi}{2N}\right)|0\rangle + \sin\left(\frac{r\pi}{2N}\right)|0\rangle, -\sin\left(\frac{r\pi}{2N}\right)|0\rangle + \cos\left(\frac{r\pi}{2N}\right)|0\rangle\right\}$, where r is an integer. For each basis,outcomes 0 and 1 are defined to correspond, respectively, to the projections onto the first and second basis elements. N and M are security parameters, that correspond to large positive integers. Their precise value (or lower bound value) depends on how secure you want the protocol to be, as we will discuss later on. The protocol then goes as follows:

1- Alice and Bob share $n = MN^2$ pairs of systems, each in the singlet state, i.e. the maximally entangled state $|\psi_-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$.

2- Alice and Bob choose independent random elements r_A^i and r_B^i of the set $\{0, \ldots, N-1\}$, for each $i \in \{1, MN^2\}$, and measure the i^{th} particle in the bases $A_i = X_{r_A^i}$ and $B_i = X_{r_B^i}$.

3- When all their measurements are complete, Alice and Bob announce their bases over a public, authenticated, classical channel.

4- First 'abort' test: Alice and Bob abort the protocol and restart unless[1]

$$2MN \leq \sum_i \sum_{c=0,\pm1} |\{j \,:\, A_j = X_i \wedge B_j = X_{i+c}\}|\,.$$

This criterion gives a lower bound on the number of runs that Alice and Bob choose neighboring or identical bases to measure their qubits. Here, two bases A_i and B_j are called 'neighboring' if $|i-j| = 1$, and identical if $i = j$.

[1] Notice that X_{r+N} contains the same basis states as X_r with the outcome conventions reversed; i.e., we interpret the bases X_{-1} and X_N to be X_{N-1} and X_0 with outcomes reversed.

5- Alice and Bob choose randomly one run among those where the chosen bases are neighboring or identical. For this run, the outcomes are kept secret (with this the secret pair of bits will be distilled later). The outcomes for all the remaining pairs (for all basis choices) are then announced.

6- Second 'abort' test: Alice and Bob abort the protocol if their outcomes a and b are not anticorrelated (i.e., $a \neq b$) in all the cases where they chose neighboring or identical bases.

7- If the protocol is not aborted, their unannounced outcomes define the secret bit, which is taken by Alice to be equal to her outcome and by Bob to be opposite to his.

To analyze the security of this protocol, we must describe formally the actions available to postquantum eavesdroppers. To give Eve maximum power, we assume that each pair of systems is produced by a source under her control. In a general, or collective, attack, Eve prepares $2n+1$ systems in a postquantum state λ sending n systems to Alice, n to Bob, and keeping 1. The state λ defines measurement probabilities $P_\lambda(abe|ABE)$, where

- $A = \{A_1, \ldots, A_n\}$ is a set of Alice's possible measurement choices, and a their joint outcome,
- $B = \{B_1, \ldots, B_n\}$ is a set of Bob's possible measurement choices, and b their joint outcome,
- $E = \{E_1\}$ a set containing a possible measurement choice of Eve, with corresponding outcome e.

This state may be nonquantum and nonlocal, but must not allow signaling even if the parties cooperate. Thus, for any partitionings $A = A^1 \cup A^2$, $B = B^1 \cup B^2$, and $E = E^1 \cup E^2$ (possibly including empty subsets), and any alternative choices $\overline{A}^2$, $\overline{B}^2$, $\overline{E}^2$ the correlations should satisfy

$$\sum_{a_2,b_2,c_2} P_\lambda(a_1a_2b_1b_2c_1c_2|A^1A^2B^1B^2E^1E^2) = \sum_{a_2,b_2,c_2} P_\lambda(a_1a_2b_1b_2c_1c_2|A^1\overline{A}^2B^1\overline{B}^2E^1\overline{E}^2).$$

Eve may wait until all Alice's and Bob's communications are finished before performing her measurement.

Now, to prove security of the protocol, the authors use the following Bell inequality:

$$t_j = \tfrac{1}{3N} \sum_{c=0,\pm1} \sum_{i=0}^{N-1} P_\lambda(a_j \neq b_j|A_j = X_i \wedge B_j = X_{i+c}) \underset{LC}{\leq} 1 - \tfrac{2}{3N}.$$

This inequality can indeed be violated by quantum mechanics whenever Alice and Bob are in the ideal scenario where no eavesdropping happens.

In the general case, it is violation of this inequality that allows Eve's knowledge to be bounded. To see this, first notice the lower bound on the value of t_s for the secret pair s:

Exercise 8.1. Given that Alice's and Bob's tests are passed and that Eve is not using a strategy that almost always fails the tests, for any λ such that $P_\lambda(\text{pass}) > \epsilon$,

$$P_\lambda(a_s \neq b_s|\text{pass}) > 1 - \frac{q}{2MN\epsilon}.$$

Hence, the no signaling condition and the chain rule for conditional probabilities, conditioned on passing the test, imply

$$t_s > 1 - \frac{1}{2MN\epsilon}. \tag{8.1}$$

Next we will see that this lower bound on t_s implies an upper bound on Eve's information, which can be made arbitrarily small as M, N become large. So let us suppose that this is actually not the case: that with probability $\delta > 0$ Eve gets an outcome e_0 such that

$$P_\lambda(a_s = b, b_s = \bar{b}|A_s = X_k, B_k = X_{k+d}, e_0) > \tfrac{1}{2}(1 + \delta')\,,$$

for some k, where $d \in \{0, \pm 1\}$, $\delta' > 0$ and $b \in \{0, 1\}$. The no signalling condition allows us to define

$$\begin{aligned} p_i^A &\equiv P_\lambda(a_s = b|A_s = X_i, e_0) \\ p_i^B &\equiv P_\lambda(b_s = \bar{b}|B_s = X_i, e_0). \end{aligned}$$

Hence, $p_k^A, p_{k+d}^B > \frac{1}{2}(1 + \delta')$.

Now,

$$\begin{aligned} P_\lambda(a_s \neq b_s|A_s = X_i, B_k = X_{i+c}, e_0) &= P_\lambda(a_s = b, b_s = \bar{b}|A_s = X_i, B_k = X_{i_c}, e_0) \\ &\quad + P_\lambda(a_s = \bar{b}, b_s = b|A_s = X_i, B_k = X_{i_c}, e_0) \\ &\leq \min(p_i^A, p_{i+c}^B) + \min(1 - p_i^A, 1 - p_{i+c}^B) \\ &= 1 - \left|p_i^A - p_{i+c}^B\right| \,. \end{aligned}$$

Hence,

$$\sum_{c=0,\pm 1} \sum_{i=0}^{N-1} P_\lambda(a_j \neq b_j|A_j = X_i \wedge B_j = X_{i+c}, e_0) \leq 3N - \left|2p_k^A - 1\right| \leq 3N - \delta'\,.$$

This implies that, conditioned on passing the test,

$$t_s \leq 1 - \frac{\delta\delta'}{3N} . \tag{8.2}$$

Now, for any fixed δ, δ', we can choose M, N, ϵ such that eq. (8.2) is incompatible with eq. (8.1), and that at the same time quantum correlations are unlikely to fail. Indeed, it suffices to take $M = N^{3/4}$ and $\epsilon = N^{-1/4}$: here, eq. (8.1) takes the form $t_s > 1 - \frac{1}{2N^{3/2}}$ and the inconsistency follows for large N.

We see then how nonlocality is crucial to the success of the protocol. If Alice and Bob were violating no Bell inequality, then Eve could eavesdrop perfectly by preparing each pair of systems in a postquantum state that is deterministic and local. This would give Eve perfect information about Alice's and Bob's measurement outcomes. On the other hand, if Alice and Bob are violating a Bell inequality, then at least some of the postquantum states prepared by Eve must be nonlocal. But any state that is deterministic and nonlocal allows signaling. So this trivial eavesdropping strategy is not available to Eve anymore. In general, the protocol works because, once the no signaling condition is assumed, nonlocal correlations must satisfy monogamy relations, hence the strength of the correlations Alice and Bob share with the eavesdropper is upper-bounded.

8.1.2 Device-independent randomness amplification and expansion

Quantum key distribution is not the only crypto task that we would like to realise in a device-independent manner. Other comprise the 'production' of randomness that is required to generate the input choices in such DIQKD tests. Technically speaking, randomness cannot actually be created. All that can be done is to take an original 'seed' of randomness and process it, to either (i) generate a new (longer) string of (less) random dits, or to (ii) generate a new (shorter) string of (more) random dits. The former task is denoted *randomness expansion*, while the latter is known as *randomness amplification*. In this section, we'll discuss both.

Randomness expansion

So let us begin with randomness expansion. This study was initiated by Colbeck [88], who noticed a relation between nonlocality and randomness which suggests to use Bell-violating devices to certify the generation of random numbers in a DI manner. A typical protocol for device-independent randomness generation (DIRNG) then uses these devices n times in succession, and generically consists of the following steps:

- A measurement step, where the successive pairs of inputs $(x_1, y_1), \ldots, (x_n, y_n)$ are used in the devices, yielding a sequence of outputs $(a_1, b_1), \ldots, (a_n, b_n)$.

- An estimation step, where the raw data is used to estimate a Bell parameter (if this parameter is too low, the protocol may abort).
- A randomness extraction step, where the raw output string is processed to obtain a smaller final string $r = r_1, \ldots, r_K$, which is uniformly random and private with respect to any potential adversary.

These protocols consume some initial random seed for choosing the inputs in the measurement step and for processing the raw data in the randomness extraction step. If more randomness is generated than is initially consumed, one has then achieved device-independent randomness expansion.

As an example, let us discuss the generic family of protocols from [89], which are based on arbitrary Bell inequalities and achieve quadratic expansion. The devices considered here have input cardinality m and output cardinality d. The first step is to find a quantitative relation between the violation of a Bell inequality and randomness. Here, the randomness of the data produced by a device, that is the randomness of the output pairs conditioned on the input pairs, is quantified by the min-Entropy

$$H_\infty(p) = -\log_2 \max_{ab} p(ab|xy) = \min_{ab}[-\log_2 p(ab|xy)]\,.$$

Now consider an arbitrary Bell functional in an $(2, m, d)$ Bell scenario

$$I(p) = \sum_{a,b,x,y} c_{a,b,x,y}\, p(ab|xy) \underset{LC}{\leq} 1\,.$$

The aim then is to find a lower bound[2] on the min-Entropy of the data in terms of its value I of the above Bell inequality, i.e.

$$H_\infty(p) \geq f(I)\,.$$

For this, consider the following optimisation problem

$$\begin{aligned} p^* \quad = \quad & \max_{a,b,x,y} \quad p(ab|xy) \\ & \quad\text{st} \quad I(p) == I \\ & \qquad\;\; p \in \mathcal{Q}\,. \end{aligned} \tag{8.3}$$

From its solution, it follows that

$$H_\infty(p) = -\log_2 \max_{ab} p(ab|xy) \geq -\log_2 p^*\,.$$

[2] Actually, they only demand such a bound to be satisfied for quantum data, i.e. quantum p.

Hence, $f(I) = -\log_2 p^*$ gives us a tight lower bound for the min-Entropy of the data. Note, however, that the optimisation problem of (8.3) involves optimising over the quantum set, which is something that cannot be done. Hence, one can effectively approximate $f(I)$ by making use of the NPA hierarchy previously discussed. By replacing the constraint $p \in \mathcal{Q}$ in problem (8.3) with $p \in \mathcal{Q}_k$, the optimisation becomes an SDP, which can be efficiently computed. Note that since $\mathcal{Q}_k$ is larger than the quantum set, the new value p_k^* obtained by solving the new version of (8.3) is potentially larger than the p^* we aim for, hence the corresponding lower bound $f_k(I) = -\log_2 p_k^*$ is smaller. The NPA hierarchy then yields a hierarchy of lower bounds to $H_\infty(p)$ that ultimately converges into $f(I)$:

$$f_1(I) \leq f_2(I) \leq \ldots \leq f_\infty(I) \equiv f(I) \leq H_\infty(p)\,.$$

The exact form of these lower bounds depends on the Bell inequality specified by the coefficients $c_{a,b,x,y}$, but they are all convex functions that are equal to 0 whenever $p \in \mathcal{C}$.

On the other hand, one can also compute lower bounds on the solution of (8.3), which translate into upper bounds on $f(I)$. To do this, one can search numerically for solutions to (8.3) with fixed Hilbert space-dimension. Specifically, one can introduce a parameterization of the state and measurement operators and vary the parameters to maximize (8.3). Whenever one finds an upper-bound of $f(I)$ this way, and notices that it coincides with $f_k(I)$ for some k, then this fixes the value of $f(I)$ and solves (8.3).

For the particular case of the CHSH inequality, [89] also finds a tight analytical lower bound to $H_\infty(p)$. To do this, they use the fact that for the case of binary measurements, the maximum value of a Bell inequality may be achieved by measuring a two-qubit system [90] on the state $|\psi_\theta\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$, for some value of θ. We leave it as an exercise to go over the derivation of such bound in [89].

Now, to use these results to lower bound the randomness of the experimental data produced by devices that violate a Bell inequality, one has first to estimate the Bell violation. This requires to use the devices a large number n of times in succession. In full generality, one cannot assume, however, that the devices behave identically and independently at each use. For instance, they may have an internal memory, so that what happens at the i^{th} use of the devices depends on what happened during the $i-1$ previous uses. The lower bounds discussed above then need to be combined with a statistical approach that takes into account such memory effects. The student who is interested in the details can consult the supplemental material of [89].

Altogether, the randomness of the output string $r = (a_1, b_1, \ldots, a_n, b_n)$ conditioned on the input string $s = (x_1, y_1, \ldots, x_n, y_n)$ is bounded by

$$H_\infty(R|S) \geq nf(\hat{I} - \epsilon) \tag{8.4}$$

with probability greater than $1-\delta$, where the statistical parameter ϵ is $\epsilon O\left(\sqrt{\frac{-\log\delta}{q^2 n}}\right)$, and $q = \min_{x,y} p(x, y)$ is the probability of the least probable input pair. The quantity $\hat{I}$ is the estimation of the value of the Bell functional from the data set obtained in the n rounds:

$$\hat{I} = \tfrac{1}{n} \sum_{a,b,x,y} c_{a,b,x,y} \frac{N(a, b, x, y)}{p(xy)},$$

where $N(a, b, x, y)$ is the number of rounds among those n where a, b were obtained and x, y were input.

Once we have obtained the string r whose entropy is bounded as in eq. (8.4), it can then be classically processed using a randomness extractor, to convert it into a string of size $nf(\hat{I} - \epsilon)$ that is nearly uniform and uncorrelated to the information of an adversary.

Randomness amplification

Let us move on to the task of randomness amplification. Here, we aim at extracting perfect (or arbitrarily close to perfect) randomness from an initial source that is partly correlated with the devices and the adversary.

The idea then is the following. Alice and Bob share a device whose statistics are compatible with quantum theory, and that violates a Bell inequality. Then, this device may allow to generate outputs that are nevertheless perfectly random, even if the inputs are not chosen perfectly at random. By this, the initial source of (imperfect) randomness that produces the measurement choices can be 'amplified' by generating (with the assistance of the device) this (more perfectly) random output string.

Before we get into the details, we need some precise notion of what *partially free randomness* is. The main idea is that, given a particular causal structure, a variable is free if it is uncorrelated with all other values except those that lie in its causal future. The results are usually independently of the exact causal structure, but it is natural to consider the one arising from relativistic spacetime, which we have used to describe Bell scenarios throughout the lectures. Given a causal structure, a variable X is perfectly free if it is uniformly distributed conditioned on any variables that cannot be caused by X. For the particular case of a relativistic understanding of "cause", X is then free if there is no reference frame in which it is correlated with variables in its past. Partial freedom may also be defined, and reads as follows: X

is ϵ-free if it is ϵ-close in variational distance to being perfectly free. This measure of distance can be operationally understood as follows: if two distributions have variational distance at most ϵ, then the probability that we ever notice a difference between them is at most ϵ.

Bell's theorem then shows that, if we have perfect free randomness to choose the measurement settings, then the violation of a Bell inequality indicates that the measurement outcomes cannot be completely pre-determined. These works on randomness amplification then extend Bell's discussion to show the following: quantum correlations can be so strong that, even if we cannot choose the measurements perfectly freely, the outputs are nevertheless perfectly free.

Pioneering work by Colbeck and Renner showed that free randomness can indeed be amplified both quantumly and in a device-independent manner [91]. For the former, i.e. by means of any [any?] box $p(ab|xy)$ that admits a quantum realisation, they showed that a source of ϵ-free bits can be used to generate arbitrarily free bits for any $\epsilon < \frac{(\sqrt{2}-1)^2}{2}$. For the latter case, i.e. by means of any $p(ab|xy)$ within the no-signalling polytope, ϵ-free bits can generate arbitrarily free bits for any $\epsilon < 0.058$.

Gallego *et al.* improved on this result by showing that the randomness of ϵ-free bits can be amplified for any value of ϵ [92]. In their seminal paper, they ask the first and most fundamental question of whether the process is at all possible, and hence restrict their analysis to the problem of generating a single final free random bit k. From a cryptographic perspective then, one must assume the worst-case scenario where all the devices we use may have been prepared by an adversary Eve equipped with arbitrary non-signalling resources, possibly even post-quantum ones. To demonstrate full randomness amplification then one must show that Eve's correlations with k can be made arbitrarily small.

An important contribution of [92] is that Bell tests for which quantum correlations achieve the maximal non-signalling violation, also known as Greenberger–Horne–Zeilinger (GHZ)-type paradoxes, are necessary for full randomness amplification. This is due to the fact that unless the maximal non-signalling violation is attained, for sufficiently small ϵ, Eve may fake the observed correlations with classical deterministic resources.The protocol that they present is then based on the violation of the five-party Mermin inequality [63].

8.1.3 Dimension witness

In these lectures we have seen how the violation of a Bell inequality can be used, from a fundamental perspective, to certify in a device independent manner the preparation of an entangled quantum state. About a decade ago, Brunner *et al.* [93] introduced another fundamental question, different in spirit to the usual classical-vs-quantum one, that can also be assessed in this black-box scenario: what is the dimension of the experimentally realised Hilbert space. In reality, the specific dimension cannot be certified, but instead one can give lower bounds to it.

The problem of testing the dimension of a Hilbert space is relevant for various reasons. On the one hand, from an information-theoretical point of view, the dimensionality of quantum systems can be seen as a resource. Thus information about Hilbert space dimension is important for quantifying the power of quantum correlations, a central issue in Quantum Information science. On the other hand, QKD security proofs rely on knowledge of the Hilbert space dimension. Even though this requirement was bypassed by the development of device-independent protocols, to prove security in traditional QKD it is useful to understand how it is possible to bound effectively the dimension of the systems distributed by the eavesdropper.

In this section we briefly review the witnesses of Brunner *et al.*, to introduce the reader to the problem of bounding the Hilbert space dimension. For a more general overview, the reader is referred to [22]. Given a correlation $p(ab|xy)$, itis said to have a d-dimensional representation if one of the following holds:

- $p(ab|xy) = \mathrm{tr}\left\{M_{a|x} \otimes M_{b|y}\,\rho\right\}$., where ρ is a state in $\mathbb{C}^d \otimes \mathbb{C}^d$, and both $M_{a|x}$ and $M_{b|y}$ are measurement operators acting on $\mathbb{C}^d$,
- $p(ab|xy)$ it can be written as a convex combination of probabilities of the form given by the previous item.

This definition may sound surprising, since we allow for convex combinations of d-dimensional correlations to qualify as d-dimensional as well. This is indeed sensible since, from a quantum information perspective, classical resources are taken to be free and we want to bound the necessary quantum resources, in this case the dimensionality of the quantum states, to achieve a task. In this scenario then shared randomness is unrestricted and the set of d-dimensional quantum correlations $\mathcal{Q}^d$ is by definition convex.

A dimension witness is then a linear functional $\vec{w}\cdot\vec{p} := \sum_{a,b,x,y} w_{abxy}\, p(ab|xy)$ such that

$$\vec{w}\cdot\vec{p} \leq w_d \quad \forall\, p \in \mathcal{Q}^d \tag{8.5}$$

and $\vec{w}\cdot\vec{p} > w_d$ for some $p \in \mathcal{Q}$.

When a correlation p violates (8.5) then it cannot be realised by measuring quantum systems of dimension at most d.

As an example, consider the asymmetric bipartite Bell scenario where Bob chooses among three dichotomic measurements and Alice between two: one dichotomic ($x = 0$) and one ternary ($x = 1$). Take then the Bell functional

$$I(p) = p_A(0|0) - p(00|00) - p(00|01) - p(00|02) + p(00|10) + p(10|11) + p(20|12) + 1\,. \tag{8.6}$$

This expression satisfies $I(p) \underset{\mathrm{LC}}{\leq} 0 \underset{\mathcal{Q}}{\leq} 0.2532$. The quantum bound is certified by the NPA hierarchy and realised by measuring a partially entangled state of two-qutrits.

Brunner *et al.* then showed that the largest violation for qubits is strictly smaller than this Tsirelson's bound. To prove this, they used the following:

- since the expression to be maximised is linear in p, the maximum will be attained by pure states $\rho = |\psi\rangle\langle\psi|$ and extremal POVMs.
- up to a local change of basis, any pure two-qubit state can be written as $|\psi(\theta)\rangle = \cos(\theta)|00\rangle + \sin(\theta)|11\rangle$.
- any extremal POVM M for qubits has elements $\{M_i\}$ which are proportional to rank 1 projectors [94] and can be parametrised via the Pauli matrices $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ as:

$$M_i = \tfrac{1}{2}\left(m_i \mathbb{1} + \vec{m}_i \cdot \vec{\sigma}\right)\,,$$

  where $\sum_i m_i = 2$, $\sum_i \vec{m}_i = 0$, $m_i^2 = \vec{m}_i^2$.

Now, optimising the linear functional over the values of the free variables that define the state $|\psi(\theta)\rangle$ and the POVMs by Alice and Bob is a not-convex quadratic program which is difficult to tackle. The authors hence opted to derive an upper bound to the maximum value of $I(p)$ using semidefinite program. This upper bound they found could moreover be realised by measuring a two-qubit state, and yields the value of 0.2071. Hence, $I(p) \underset{\mathcal{Q}^2}{\leq} 0.2071 \underset{\mathcal{Q}}{\leq} 0.2532$.

Hence, we see how linear functionals on the elements of the probability vector can be used to prove the dimensionality of the Hilbert space of an experimental implementation.

Bibliography

[1] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

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

[3] H. M. Wiseman. The two bell's theorems of john bell. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424001, 2014.

[4] John S. Bell. The theory of local beables. *Epistemological Letters*, 9:11–24, 1976.

[5] Arthur Fine. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.*, 48(5):291–295, 1982.

[6] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23(15):880–884, Oct 1969.

[7] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 2003.

[8] John Preskill. Lecture Notes for Physics 229: Quantum Information and Computation, 1998. http://www.theory.caltech.edu/people/preskill/ph219/.

[9] M. A. Naimark. On a representation of additive operator valued set functions (Russian). *Doklady Acad. Nauk SSSR*, 41(5):373–375, 1943.

[10] Vern Paulsen. Lecture Notes for Entanglement and Non-Locality, 2016.

[11] TOBIAS FRITZ. Tsirelson's problem and kirchberg's conjecture. *Reviews in Mathematical Physics*, 24(05):1250012, 2012.

[12] Vern Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge University Press, 2003.

[13] B. S. Tsirelson. Bell inequalities and operator algebras, problem statement for website of open problems, 2006. `https://qig.itp.uni-hannover.de/qiproblems/Bell_inequalities_and_operator_algebras`.

[14] M. Navascués, T. Cooney, D. Pérez-García, and N. Villanueva. A Physical Approach to Tsirelson's Problem. *Foundations of Physics*, 42(8):985–995, 2012.

[15] William Slofstra. Tsirelson's problem and an embedding theorem for groups arising from non-local games, 2016. Preprint at arxiv:1606.03140.

[16] N. Gisin. Bell's inequality holds for all non-product states. *Physics Letters A*, 154(5):201 – 202, 1991.

[17] Reinhard F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, Oct 1989.

[18] Sandu Popescu. Bell's inequalities versus teleportation: What is nonlocality? *Phys. Rev. Lett.*, 72:797–799, Feb 1994.

[19] Jonathan Barrett. Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a bell inequality. *Phys. Rev. A*, 65:042302, Mar 2002.

[20] Antonio Acín, Nicolas Gisin, and Benjamin Toner. Grothendieck's constant and local models for noisy entangled quantum states. *Phys. Rev. A*, 73:062105, Jun 2006.

[21] Sandu Popescu. Bell's inequalities and density matrices: Revealing "hidden" nonlocality. *Phys. Rev. Lett.*, 74:2619–2622, Apr 1995.

[22] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419–478, Apr 2014.

[23] Yeong-Cherng Liang and Andrew C. Doherty. Better bell-inequality violation by collective measurements. *Phys. Rev. A*, 73:052116, May 2006.

[24] Miguel Navascués and Tamás Vértesi. Activation of nonlocal quantum resources. *Phys. Rev. Lett.*, 106:060403, Feb 2011.

[25] Carlos Palazuelos. Superactivation of quantum nonlocality. *Phys. Rev. Lett.*, 109:190401, Nov 2012.

[26] Daniel Cavalcanti, Mafalda L. Almeida, Valerio Scarani, and Antonio Acín. Quantum networks reveal quantum nonlocality. *Nature Communications*, 2:184, 2011.

[27] Flavien Hirsch, Marco Túlio Quintino, Joseph Bowles, and Nicolas Brunner. Genuine hidden quantum nonlocality. *Phys. Rev. Lett.*, 111:160402, Oct 2013.

[28] Stuart J. Freedman and John F. Clauser. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941, Apr 1972.

[29] Alain Aspect, Philippe Grangier, and Gérard Roger. Experimental tests of realistic local theories via bell's theorem. *Phys. Rev. Lett.*, 47:460–463, Aug 1981.

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

[31] Marco Genovese. Research on hidden variable theories: A review of recent progresses. *Physics Reports*, 413(6):319 – 396, 2005.

[32] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. Experimental violation of a Bell's inequality with efficient detection. *Nature*, 409:791, 2001.

[33] Julian Hofmann, Michael Krug, Norbert Ortegel, Lea Gérard, Markus Weber, Wenjamin Rosenfeld, and Harald Weinfurter. Heralded entanglement between widely separated atoms. *Science*, 337(6090):72–75, 2012.

[34] Richard D. Gill. Statistics, causality and bell's theorem. *Statist. Sci.*, 29(4):512–528, 11 2014.

[35] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682 – 686.

[36] Marissa Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K.

Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger. Significant-loophole-free test of bell's theorem with entangled photons. *Phys. Rev. Lett.*, 115:250401, Dec 2015.

[37] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, Scott Glancy, Deny R. Hamel, Michael S. Allman, Kevin J. Coakley, Shellee D. Dyer, Carson Hodge, Adriana E. Lita, Varun B. Verma, Camilla Lambrocco, Edward Tortorici, Alan L. Migdall, Yanbao Zhang, Daniel R. Kumor, William H. Farr, Francesco Marsili, Matthew D. Shaw, Jeffrey A. Stern, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Thomas Jennewein, Morgan W. Mitchell, Paul G. Kwiat, Joshua C. Bienfang, Richard P. Mirin, Emanuel Knill, and Sae Woo Nam. Strong loophole-free test of local realism. *Phys. Rev. Lett.*, 115:250402, Dec 2015.

[38] B.S. Tsirelson. Some results and problems on quantum Bell-type inequalities. *Hadronic Journal Supplement*, 8:329–345, 1993.

[39] Daniel Collins and Nicolas Gisin. A relevant two qubit bell inequality inequivalent to the chsh inequality. *Journal of Physics A: Mathematical and General*, 37(5):1775, 2004.

[40] David Avis, Hiroshi Imai, Tsuyoshi Ito, and Yuuya Sasaki. Deriving Tight Bell Inequalities for 2 Parties with Many 2-valued Observables from Facets of Cut Polytopes, 2004. Preprint at quant-ph/0404014.

[41] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger. Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial. *Phys. Rev. Lett.*, 96:250401, Jun 2006.

[42] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, October 2009.

[43] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum nonlocality and beyond: Limits from nonlocal computation. *Phys. Rev. Lett.*, 99:180502, Oct 2007.

[44] M. Navascués and H. Wunderlich. A glance beyond the quantum model. *Proc. R. Soc. A*, 466(2115):881–890, 2010.

[45] Joe Henson and Ana Belén Sainz. Macroscopic noncontextuality as a principle for almost-quantum correlations. *Phys. Rev. A*, 91:042114, Apr 2015.

[46] T. Fritz, A.B. Sainz, R. Augusiak, J Bohr Brask, R. Chaves, A. Leverrier, and A. Acín. Local orthogonality as a multipartite principle for quantum correlations. *Nature Communications*, 4:2263, 2013.

[47] A. B. Sainz, T. Fritz, R. Augusiak, J. Bohr Brask, R. Chaves, A. Leverrier, and A. Acín. Exploring the local orthogonality principle. *Phys. Rev. A*, 89:032117, Mar 2014.

[48] 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.

[49] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98(1):010401, 2007.

[50] Stefano Pironio, Miguel Navascués, and Antonio Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM Journal on Optimization*, 20(5):2157–2180, 2010.

[51] Károly F. Pál and Tamás Vértesi. Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems. *Phys. Rev. A*, 82:022116, Aug 2010.

[52] Miguel Navascués, Yelena Guryanova, Matty J. Hoban, and Antonio Acín. Almost quantum correlations. *Nature Communications*, 6:6288, 2015.

[53] George Svetlichny. Distinguishing three-body from two-body nonseparability by a bell-type inequality. *Phys. Rev. D*, 35:3066–3069, May 1987.

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

[55] Jean-Daniel Bancal, Jonathan Barrett, Nicolas Gisin, and Stefano Pironio. Definitions of multipartite nonlocality. *Phys. Rev. A*, 88:014102, Jul 2013.

[56] Rodrigo Gallego, Lars Erik Würflinger, Antonio Acín, and Miguel Navascués. Operational framework for nonlocality. *Phys. Rev. Lett.*, 109:070401, Aug 2012.

[57] Benjamin Toner and Frank Verstraete. Monogamy of Bell correlations and Tsirelson's bound, 2006. Preprint at arXiv:quant-ph/0611001.

[58] Ll. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73:012112, Jan 2006.

[59] Marcin Pawłowski and Časlav Brukner. Monogamy of bell's inequality violations in nonsignaling theories. *Phys. Rev. Lett.*, 102:030403, Jan 2009.

[60] Simon Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *J. Math. Mech.*, 17:59–87, 1967.

[61] Robert W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A*, 71(5):052108, 2005.

[62] Alexander A. Klyachko, M. Ali Can, Sinem Binicioğlu, and Alexander S. Shumovsky. Simple test for hidden variables in spin-1 systems. *Phys. Rev. Lett.*, 101:020403, Jul 2008.

[63] N. David Mermin. Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.*, 65:3373–3376, Dec 1990.

[64] Asher Peres. Incompatible results of quantum measurements. *Physics Letters A*, 151(3):107 – 108, 1990.

[65] A. Cabello, S. Severini, and A. Winter. (Non-)Contextuality of Physical Theories as an Axiom. October 2010. Preprint at http://arXiv.org/abs/1010.2163.

[66] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, 2011.

[67] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, cohomology and paradox. 24^{th} *Annual Conference on Computer Science Logic*, pages 211–228, 2015.

[68] Lucien Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Phys. Rev. Lett.*, 71:1665–1668, Sep 1993.

[69] Antonio Acín, Tobias Fritz, Anthony Leverrier, and Ana Belén Sainz. A combinatorial approach to nonlocality and contextuality. *Communications in Mathematical Physics*, 334(2):533–628, 2015.

[70] Adán Cabello, José M. Estebaranz, and Guillermo García-Alcaine. Bell-Kochen-Specker theorem: a proof with 18 vectors. *Phys. Lett. A*, 212(4):183–187, 1996.

[71] Adán Cabello. Experimentally testable state-independent quantum contextuality. *Phys. Rev. Lett.*, 101(21):210401, 2008.

[72] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[73] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[74] K. C. Toh, M.J. Todd, and R. H. Tütüncü. Sdpt3 – a matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.

[75] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using sdpt3. *Mathematical Programming*, 95(2):189–217, Feb 2003.

[76] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999. Version 1.05 available from `http://fewcal.kub.nl/sturm`.

[77] Thomas Christof, Andreas Löbel, and Sebastian Schenker. POlyhedron Representation Transformation Algorithm, 1997. http://www.iwr.uni-heidelberg.de/groups/comopt/software/PORTA/.

[78] Komei Fukuda. cdd/cdd+ reference manual. ETH Zurich, Switzerland, 1997.

[79] Stefan Lörwald and Gerhard Reinelt. Panda: a software for polyhedral transformations. *EURO Journal on Computational Optimization*, pages 1–12, 2015.

[80] Cyril Branciard. Detection loophole in bell experiments: How postselection modifies the requirements to observe nonlocality. *Phys. Rev. A*, 83:032123, Mar 2011.

[81] Philippe H. Eberhard. Background level and counter efficiencies required for a loophole-free einstein-podolsky-rosen experiment. *Phys. Rev. A*, 47:R747–R750, Feb 1993.

[82] J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein, and A. Acín. Detecting nonlocality in many-body quantum states. *Science*, 344(6189):1256–1258, 2014.

[83] J. Tura, A. B. Sainz, T. Vértesi, A. Acín, M. Lewenstein, and R. Augusiak. Translationally invariant multipartite bell inequalities involving only

two-body correlators. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424024, 2014.

[84] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991.

[85] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984.

[86] Antonio Acín, Nicolas Gisin, and Lluis Masanes. From bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.*, 97:120405, Sep 2006.

[87] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95:010503, Jun 2005.

[88] Roger Colbeck. *Quantum and relativistic protocols for secure multi-party computation*. PhD thesis, University of Cambridge, 2007.

[89] 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*, 464:1021, 2010.

[90] Lluis Masanes. Extremal quantum correlations for N parties with two dichotomic observables per site. 2005. Preprint at http://arxiv.org/abs/quant-ph/0512100.

[91] Roger Colbeck and Renato Renner. Free randomness can be amplified. *Nature Physics*, 8:450–453, 2012.

[92] Rodrigo Gallego, Lluis Masanes, Gonzalo De La Torre, Chirag Dhara, Leandro Aolita, and Antonio Acín. Full randomness from arbitrarily deterministic events. *Nature Communications*, 4:2654, 2013.

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

[94] Giacomo Mauro D'Ariano, Paoloplacido Lo Presti, and Paolo Perinotti. Classical randomness in quantum measurements. *Journal of Physics A: Mathematical and General*, 38(26):5979, 2005.