Limits on Nonlocal Correlations from Physical Principles: A Review

Ravi Kunjwal

The Institute of Mathematical Sciences

CIT Campus, Taramani, Chennai 600 113, India

Thesis submitted in partial fulfillment of the requirements for the award of the degree of

Master of Science

HOMI BHABHA NATIONAL INSTITUTE

June 2012

BONAFIDE CERTIFICATE

Certified that this dissertation titled **Limits on Nonlocal Correlations from Physical Principles: A Review** is the bonafide work of Ravi Kunjwal who carried out the project under my supervision.

Dr. Sibasish Ghosh
Theoretical Physics
The Institute of Mathematical Sciences,
Chennai

Date

Abstract

The existence of nonlocal correlations is a feature of quantum theory that sets it apart from any conceivable classical theory obeying relativistic causality. Bell's theorem quantifies this distinction between local and nonlocal correlations using the notion of Bell inequalities. While being consistent with the premise of special relativity that no signal can travel arbitrarily fast, nonlocal correlations offer an information-theoretic advantage over correlations obtainable classically by shared randomness. The strength of nonlocal correlations allowed by quantum theory, however, is limited. There is no obvious physical principle (akin to the no-signalling principle from special relativity) that restricts the strength of quantum correlations. Indeed, there exist no-signalling correlations that are maximally nonlocal, unlike quantum correlations which are limited by Tsirelson's bound. Even though Tsirelson's bound follows from the Hilbert space structure of quantum theory, it is not clear whether there exist physical constraints that impose such a limitation on quantum correlations. It is hard to extract such a physical principle from the standard Hilbert space formalism of quantum theory. Instead, one can study correlations from the perspective of information theory and look at quantum correlations "from the outside". This approach has yielded some insights into physical and informational primitives that are obeyed by quantum correlations.

We will review recent research in this direction, seeking a better understanding of the physical principles underlying quantum theory from the vantage point of correlations. Our review includes information causality [8], macroscopic locality [18], and a relationship between the uncertainty principle and nonlocality [19].

Acknowledgements

I am grateful to Dr. Sibasish Ghosh for agreeing to supervise my M.Sc. thesis and for his patience, time, and willingness to listen to (and answer) my naive questions during the process. I would like to thank Rajarshi, for conversations over coffee, lunch or dinner, about things quantum and otherwise. Thanks are also due to Prabha, for conversations and feedback on things I mess up and things I don't quite. The weekly journal club meetings have been stimulating, and I thank Prof. R. Simon for his insightful remarks during these meetings, including the one where I gave my first talk. I am thankful to my office-mates for not minding my idiosyncrasies, especially the ever-helpful Dibyakrupa for his help with LaTeX while writing this thesis.

Finally, I am indebted to my parents and my sister for their love and support, and for giving me the freedom to follow my flights of fancy.

Contents

List of Figures

List of Tables

Chapter 1

Bell nonlocality and Quantum Theory

1.1 Introduction

There are many reasons why quantum theory is a drastic departure from classical ways of thinking about the world. This nonclassicality of quantum theory can be attributed to features like the noncommutativity of observables, entanglement, and the intrinsically probabilistic nature of predictions via the Born rule. The feature that this thesis will focus on is nonlocality (or more precisely, nonlocal correlations), as defined by John Bell in his seminal work [1, 2]. This chapter is an introduction to Bell's theorem, the associated notion of Bell Inequalities and their violations allowed by quantum theory. Violation of a Bell Inequality is a signature of nonlocal correlations.

1.2 Bell's Theorem

The mathematical formalism of quantum theory (QT) gives rise to all its counterintuitive consequences. Of all these consequences, however, possibly the most intriguing one that separates quantum theory from any conceivable classical theory is the notion of Bell nonlocality. Bell's notion of locality is formulated in terms of constraints that a joint probability distribution arising from a Bell test in any classical theory must satisfy. By 'classical', we refer to local hidden variable (LHV) models that seek to recover quantum theory as an approximation to some deeper classical theory, just as classical statistical mechanics is considered a statistical approximation to the exact classical mechanics of many particles. In other words, LHV models seek to "complete" quantum theory by supplementing it with hidden variables ('hidden' because they are not accessible to experiments). Bell's theorem shows why one can rule out any LHV description of QT – an LHV model must satisfy Bell inequalities, while QT allows one to violate these inequalities.

1.2.1 The Bell scenario

In the most general correlation experiment, a system (consisting of N subsystems) is prepared in some state. After state-preparation, each of the N subsystems is sent to a different party. Each party can then perform measurements on the subsystem in its laboratory. We consider the case of N parties, M measurement choices per party, and K outcomes per measurement choice. This is the (N,M,K) Bell scenario. We also ensure that the N parties are spacelike separated during the experiment – this rules out the possibility of any signalling between the parties while the experiment is performed. In each run of the experiment, each party randomly measures one of the M observables and notes down which of the K possible outcomes occurs. After a large number of repetitions of the experiment, the statistics of the experiment can be written down as

$$P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N)$$
$$\text{where} \quad A_i \in \{1, ..., K\}, \quad X_i \in \{1, ..., M\}, \quad \forall i \in \{1, ...N\}$$

This correlation experiment is called a Bell test. See figure (1.1) for a schematic representation of a Bell test. Note that the choice of measurement settings is random: each party chooses any one of the M available measurements with probability $\frac{1}{M}$. This ensures measurement independence. The resulting joint probability distribution above characterizes the correlations between the N parties. The key question here is whether this probability distribution admits a local hidden variable (LHV) model, i.e., whether each (local) outcome depends only on the choice of (local) measurement, and possibly the value of some shared (hidden) variable λ which takes values according to some distribution $P(\lambda)$. The probability distribution admits an LHV model if and only if one can write it as

$$\begin{aligned} &P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N) \\ &= \sum_{\lambda} P(\lambda)P(A_1|X_1, \lambda)P(A_2|X_2, \lambda), ..., P(A_N|X_N, \lambda) \end{aligned} \tag{1.1}$$

where $P(A_i|X_i, \lambda), \forall i \in \{1, 2, ..., N\}$, is the probability that the ith party

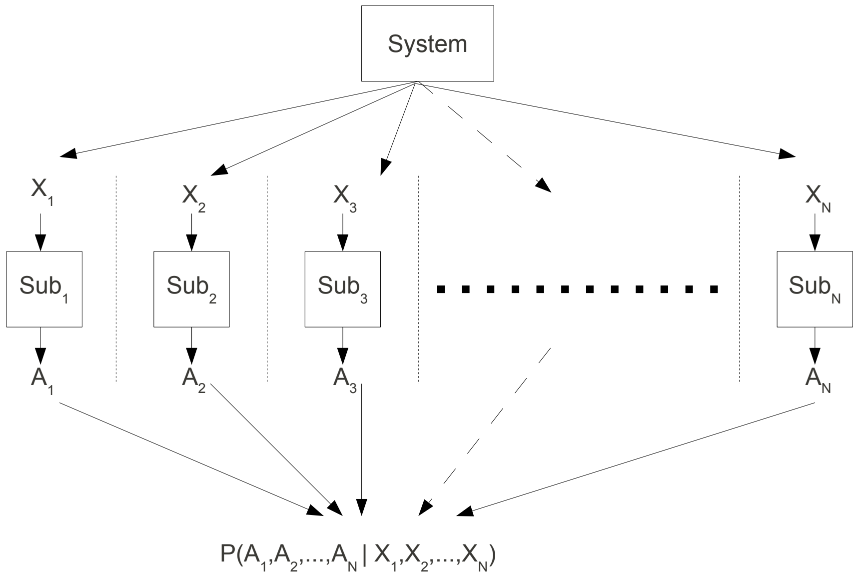

Figure 1.1: Bell test: Each of the N spacelike separated parties receives a subsystem and performs measurements on it.

obtains outcome $A_i \in \{1, ..., K\}$ for measurement $X_i \in \{1, ..., M\}$, given the shared random variable λ. We will call such a distribution 'Bell local', indicating that it admits an LHV model. Of course, λ could be a continuous parameter in which case the discrete sum above would be replaced by an integral over λ according to some probability density $P(\lambda)$. This will not affect Bell's theorem.

If, on the other hand, one cannot write the joint distribution as arising from an LHV model, the distribution is 'Bell nonlocal'. To test whether a distribution belongs to the set of Bell-local distributions, one invokes the notion of Bell Inequalities. Before we get to Bell inequalities, though, we need to introduce the notion of a Bell polytope.

1.2.2 The Bell polytope

One can write down a probability vector $\vec{P}$ of conditional probabilities

$$P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N)$$

with each entry corresponding to a particular instance of

$$(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N).$$

There are M^N choices of measurement settings $(X_1, X_2, ..., X_N)$ and for each of these settings there are K^N possible outcomes $(A_1, A_2, ..., A_N)$. So we have a total of $(MK)^N$ entries in our probability vector $\vec{P}$. This probability vector should satisfy positivity – every entry lies between 0 and 1 – and normalization, i.e.,

$$\sum_{(A_1,A_2,...,A_N)} P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N) = 1, \quad \forall (X_1, X_2, ..., X_N) \quad (1.2)$$

This is a total of M^N linear constraints and leaves us with $D = M^N(K^N - 1)$ independent probabilities in $\vec{P}$. So the probability vector $\vec{P}$ lives in $\mathbb{R}^D$ real space. The normalization and positivity constraints mean that all vectors $\vec{P}$ consistent with these belong to a subset of $\mathbb{R}^D$. This subset is a convex polytope. We will define a convex polytope when we discuss Bell Inequalities – that is where this notion becomes relevant. For now, one can treat 'convex polytope' as a label for the subset of $\mathbb{R}^D$ in which $\vec{P}$ lives.

Non-signalling (NS) Polytope:

From the requirement of spacelike separation between the N parties during the experiment, one gets 'no-signalling' constraints that $\vec{P}$ must obey. Consider any subset $\aleph$ of the N parties $\{1, 2, ..., N\}$, $\aleph = \{i_1, i_2, ..., i_n\}$ where $i_k \in \{1, 2, ..., N\} \quad \forall k \in \{1, 2, ..., n\}, \quad n \leq N$. Consider also the complement of $\aleph$, $\aleph^c$, such that $\aleph \cup \aleph^c = \{1, 2, ..., N\}$. For notational convenience, let's label the set of measurement outcomes for N parties by $A := (A_1, A_2, ..., A_N)$, the measurement settings by $X := (X_1, X_2, ..., X_N)$, measurement outcomes for parties in $\aleph$ by $A_\aleph := (A_{i_1}, A_{i_2}, ..., A_{i_n})$ and their measurement settings by $X_\aleph := (X_{i_1}, X_{i_2}, ..., X_{i_n})$, and similarly the measurement outcomes and settings for those in $\aleph^c$ are denoted by $A_{\aleph^c}$ and $X_{\aleph^c}$ respectively.

No-signalling means that for any $\aleph$ if we sum $P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N)$ over all possible outcomes for parties in $\aleph$, given a choice of measurement settings $X_\aleph$ and $X_{\aleph^c}$, the resultant will be independent of the particular choice of $X_\aleph$, i.e., we get the same answer for all choices of $X_\aleph$, keeping $X_{\aleph^c}$ fixed. This sum is the marginal conditional probability that parties in $\aleph^c$ get outcomes $A_{\aleph^c}$, given measurments $X_{\aleph_c}$. So information about the measurement settings of parties in $\aleph$ cannot be inferred by those in $\aleph^c$, since they can only observe the marginal conditional probability (which does not distinguish between different choices of $X_\aleph$). We can now state the no-signalling condition formally as:

$$\sum_{A_\aleph} P(A|X) = \sum_{A_\aleph} P(A|X'), \tag{1.3}$$

$\forall X \neq X'$ such that $X_\aleph \neq X'_\aleph$ and $X_{\aleph^c} = X'_{\aleph^c}$, for any choice of $\aleph \subseteq \{1, 2, ..., N\}$

This quantity defines the marginal:

$$P(A_{\aleph^c}|X_{\aleph^c}) := \sum_{A_\aleph} P(A|X) \tag{1.4}$$

This marginal is clearly independent of the choice of measurement settings $X_\aleph$. The three constraints of positivity, normalization, and no-signalling thus define a convex polytope that we will call the **Non-signalling (NS) polytope**.

Bell Polytope:

Further constraints – that of locality and realism – define the set of Bell-local probability vectors $\vec{P}$. They form a convex polytope called the Bell polytope corresponding to the given Bell scenario (N, M, K). Before we formally define what a convex polytope is, let us look at what 'locality' and 'realism' mean in the context of Bell's theorem [2].

Realism is the assumption that there exist objective properties of the system prepared in a particular state (before subsystems are sent to the N parties) and that these properties can influence measurement outcomes on each of the N subsystems once they are sent to the respective parties and the measurements performed. These properties are denoted by some ("hidden") variable $\lambda \in \Lambda$ (where Λ is the space that λ takes values in according to some probability distribution $P(\lambda)$). This $P(\lambda)$ expresses a 'coarse graining' over exact states λ in a manner that is similar to how distributions over a phase space (of exact states, e.g., (x, p)) arise in statistical mechanics – from our ignorance about the exact state.

Locality is the assumption that measurement outcomes on each subsystem depend only on the value of λ and the measurement performed. This is ensured by keeping spacelike separation between the N distinct measurement events so that there is no communication between parties during the experiment.

One can imagine λ as a list (in general, inaccessible to the experimenter, i.e., "hidden") that "instructs" each party's subsystem to behave in some specific way (i.e., produce a set of $P(A_i|X_i, \lambda)$s) when measurements are made on it. Note that the form of a Bell-local distribution (as defined in 'The Bell scenario'):

$$P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N)$$
$$= \sum_\lambda P(\lambda)P(A_1|X_1, \lambda)P(A_2|X_2, \lambda), ..., P(A_N|X_N, \lambda)$$

satisfies the no-signalling constraint. Further, it assumes local realism, since the conditional probability factors on the right-hand-side are of the form $P(A_i|X_i, \lambda)$. It turns out that the distributions allowed in an LHV model do not encompass all the possible distributions that are observed experimentally – that is where quantum theory comes in. This is essentially the statement of Bell's theorem – because quantum theory can generate Bell-nonlocal distributions, it is incompatible with local realism. No local hidden variable λ can produce Bell-nonlocal distributions. To demonstrate this, we need the notion of Bell Inequalities.

1.2.3 Bell Inequalities

To introduce Bell Inequalities, we need to formally define a convex polytope which is the geometry of the **NS polytope** and the **Bell polytope**. The definitions that follow are sufficient for our purposes. For a more detailed treatment one may refer to [3, 4]. Our discussion here also draws upon [5].

Convex set: Consider the real vector space $\mathbb{R}^D$. A convex subset, $\mathcal{C}' \subseteq \mathbb{R}^D$, is a collection of points that is closed under convex combinations, i.e., for any $x, y \in \mathcal{C}'$, we have $\mu x + (1 - \mu)y \in \mathcal{C}'$, where $0 \leq \mu \leq 1$. In more geometric terms, for any two points in the convex set $\mathcal{C}'$, the line segment joining them is also in $\mathcal{C}'$ (cf. Fig. 1.2). The intersection of any collection of convex sets is a convex set.

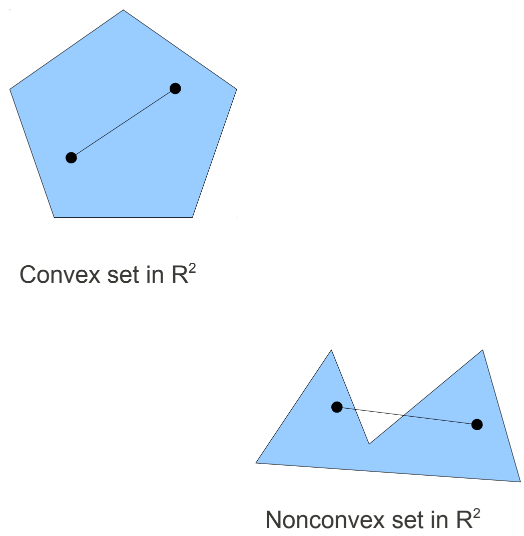

Figure 1.2: Examples of convex and nonconvex sets in $\mathbb{R}^2$

Convex hull: For any subset $K \subseteq \mathbb{R}^D$, the smallest convex set, $conv(K)$, such that $K \subseteq conv(K)$, is called the **convex hull of** K:

$$conv(K) := \bigcap \{K' \subseteq \mathbb{R}^D : K \subseteq K', K' \text{ convex}\}$$

For any finite set of points $\{x_1, ..., x_k\} \subseteq K$ and real numbers $\mu_1, ..., \mu_k \in [0, 1]$ with $\mu_1 + ... + \mu_k = 1$, we must have $\mu_1 x_1 + ... + \mu_k x_k \in conv(K)$.

Scalar Product: We define the following scalar product between any $x, y \in \mathbb{R}^D$:

$$\langle x, y \rangle = \sum_{i=1}^{D} x_i y_i$$

Hyperplane: is a set

$$H = \{x \in \mathbb{R}^D | \langle x, y \rangle = \alpha\} \text{ for some } y \in \mathbb{R}^D, y \neq 0, \text{ and some } \alpha \in \mathbb{R}$$

This hyperplane H divides $\mathbb{R}^D$ into halfspaces:

1. **Open halfspace:**

$$\{x \in \mathbb{R}^D | \langle x, y \rangle < \alpha\} \quad (\text{and } \{x \in \mathbb{R}^D | \langle x, y \rangle > \alpha\})$$

2. **Closed halfspace:**

$$\{x \in \mathbb{R}^D | \langle x, y \rangle \leq \alpha\} \quad (\text{and } \{x \in \mathbb{R}^D | \langle x, y \rangle \geq \alpha\})$$

Convex Polytope: There are two equivalent characterizations of a convex polytope: the half-space representation, and the vertex representation.

Definition 1. *(Halfspace representation)*

A convex polytope $\mathcal{C} \subseteq \mathbb{R}^D$ is the intersection of finitely many closed halfspaces.

The hyperplanes defining these closed halfspaces are called facets of the polytope and the linear inequalities that specify these closed halfspaces are called facet-defining inequalities. They form the boundary of the convex polytope $\mathcal{C}$. The other characterization of a convex polytope – in terms of points – is the vertex representation which is dual to the halfspace representation.

Definition 2. *(Vertex representation)*

A convex polytope $\mathcal{C} \subseteq \mathbb{R}^D$ is the convex hull of a finite number (E) of extremal points $v_e \in \mathbb{R}^D, e \in \{1, 2, ..., E\}$.

That is, any point in the convex polytope $\mathcal{C}$ can be written as a convex combination of the E extremal points $v_e \in \mathbb{R}^D, e \in \{1, 2, ..., E\}$.

The Non-signalling (NS) Polytope is the intersection of closed halfspaces defined by the following conditions on $\vec{P} \in \mathbb{R}^D$ (as stated earlier in Section 1.2.2):

Positivity: $0 \leq P(A|X) \leq 1, \forall A, X$

Normalization: $\sum_A P(A|X) = 1, \forall X$ (cf. eqn 1.2)

No-signalling: $\sum_{A_\aleph} P(A|X) = \sum_{A_\aleph} P(A|X') = P(A_{\aleph^c}|X_{\aleph^c}) \quad \forall X \neq X'$ (cf. eqn 1.3 and 1.4)

The Bell Polytope. A probability vector $\vec{P}$ in the Bell polytope is 'Bell local' and its elements can therefore be written as:

$$P(A_1, A_2, ..., A_N|X_1, X_2, ..., X_N)$$
$$= \sum_\lambda P(\lambda)P(A_1|X_1, \lambda)P(A_2|X_2, \lambda), ..., P(A_N|X_N, \lambda) \qquad \text{(cf. 1.1)}$$

for some $\lambda \in \Lambda$ distributed according to $P(\lambda)$, where $P(\lambda) \geq 0 \quad \forall \lambda \in \Lambda$ and $\sum_{\lambda\in\Lambda} P(\lambda) = 1$. One can rewrite equation (1.1) as a convex combination of deterministic behaviours and show how the vertex representation comes in handy in defining Bell Polytopes. Following [5], we can think of deterministic single-site maps G_i assigning outcomes to measurements on the ith subsystem:

$$G_i : \{1, 2, ..., M\} \to \{1, 2, ..., K\}$$

That is, $G_i(X_i) = A_i$, where $X_i \in \{1, ..., M\}$ and $A_i \in \{1, ..., K\}$. Each X_i is mapped to a unique A_i for the ith subsystem, so G_i is a particular deterministic map from measurements to outcomes for the ith subsystem. We can write the single-site probability

$$P(A_i|X_i, \lambda) = \sum_{G_i} P^\lambda_{G_i} \delta^{A_i}_{G_i(X_i)}$$

$P^\lambda_{G_i}$ is the probability that a particular single-site map G_i is applied on the ith subsystem, given λ. Obviously, $P^\lambda_{G_i} \geq 0$ and $\sum_{G_i} P^\lambda_{G_i} = 1$. Also, $\delta^{A_i}_{G_i(X_i)} = 1$ if and only if $G_i(X_i) = A_i$, otherwise $\delta^{A_i}_{G_i(X_i)} = 0$. The number of possible deterministic single-site maps G_i is K^M, since there are K possible outcomes for each of the M measurement choices at the ith site. Considering the

N deterministic single-site maps $(G_1, G_2, ..., G_N)$ from $(X_1, X_2, ..., X_N) \rightarrow (A_1, A_2, ..., A_N)$, one can rewrite (1.1) as:

$$P(A_1, ..., A_N|X_1, ..., X_N) = \sum_{G_1,...,G_N} p_{G_1,...,G_N} \prod_{j=1}^{N} \delta^{A_j}_{G_j(X_j)} \tag{1.5}$$

where

$$p_{G_1,...,G_N} = \sum_{\lambda} P(\lambda) \prod_{i=1}^{N} P^{\lambda}_{G_i} \tag{1.6}$$

Again, $p_{G_1,...,G_N} \geq 0$ and $\sum_{G_1,...,G_N} p_{G_1,...,G_N} = 1$. The distribution $p_{G_1,...,G_N}$ is specified by the hidden variable λ. Equation (1.5) expresses $P(A|X)$ as a convex combination of deterministic behaviours $\prod_{j=1}^{N} \delta^{A_j}_{G_j(X_j)}$.

The deterministic probabilities $P(A|X) = \prod_{j=1}^{N} \delta^{A_j}_{G_j(X_j)}$, corresponding to different choices of single-site maps $(G_1, G_2, ..., G_N)$, define the vertices of the Bell Polytope. Clearly, the Bell Polytope is the convex hull of these vertices. These vertices are also called extremal points of the Bell Polytope. We know that there are K^M single-site maps for each party. For the Bell scenario with N parties, therefore, we have K^{MN} extremal points or deterministic behaviours.

On the other hand, the facets of the Bell Polytope are characterized by facet-defining Bell Inequalities, or simply **facet Bell Inequalities**. We denote the Bell Polytope by $\mathcal{C}_B$. A facet-defining inequality can be defined as follows:

Facet-defining inequality: A facet-defining inequality for a convex polytope $\mathcal{C} \subset \mathbb{R}^D$ is one which is saturated by at least D affinely independent extreme points of $\mathcal{C}$, i.e., at least D such extreme points of $\mathcal{C}$ attain the upper bound of the inequality.

Affine independence: A set $X = \{x_1, x_2, ..., x_k\}$ is affinely independent if and only if for every $x_j \in X$ the $k-1$ points in the set $\{x_i - x_j | x_i \neq x_j\}$ are linearly independent.

The probability vectors within the Bell Polytope, $\vec{P} \in \mathcal{C}_B$, satisfy linear inequalities of the form:

$$\langle \vec{F}, \vec{P} \rangle \leq \beta_{LHV}, \quad \forall \quad \vec{P} \in \mathcal{C}_B \tag{1.7}$$

where

$$\langle \vec{F}, \vec{P} \rangle := \sum_{A,X} F_{A,X} P(A|X) \tag{1.8}$$

Clearly, equation (1.7) represents a halfspace defined by the vector $\vec{F} \in \mathbb{R}^D$ for probability vectors $\vec{P} \in \mathcal{C}_B$. The hyperplane defining this halfspace is obviously $\langle \vec{F}, \vec{P} \rangle = \beta_{LHV}$. Linear inequalities of the form (1.7) are called Bell Inequalities. They are satisfied by all the points in $\mathcal{C}_B$. A Bell Inequality may therefore be defined as follows:

Bell Inequality: A Bell Inequality is an inequality of the form (1.7) that is satisfied by all the points, $\vec{P}$, in the Bell Polytope $\mathcal{C}_B$. Also, at least one $\vec{P} \in \mathcal{C}_B$ should lie on the hyperplane, i.e., satisfy the equality in (1.7). Thus, a Bell Inequality defines a halfspace in $\mathbb{R}^D$ that divides $\mathbb{R}^D$ into two parts – one that contains $\mathcal{C}_B$ and satisfies (also saturates) the inequality, and the other that doesn't contain any point $\vec{P} \in \mathcal{C}_B$ and $\therefore$ violates the Bell Inequality.

If the hyperplane corresponding to a Bell Inequality happens to be a facet of the Bell Polytope, $\mathcal{C}_B$, then the Bell Inequality is a **facet-defining Bell Inequality**. Formally:

Facet Bell Inequality: For (1.7) to be a facet Bell Inequality, the following must hold:

$$\langle \vec{F}, \vec{P}_{G_1,\ldots,G_N} \rangle = \beta_{LHV} \tag{1.9}$$

for at least D affinely independent extremal points (vertices) $\vec{P}_{G_1,\ldots,G_N}$ of $\mathcal{C}_B$ with elements given by $P_{G_1,\ldots,G_N}(A_1, \ldots, A_N|X_1, \ldots, X_N) = \prod_{j=1}^{N} \delta^{A_j}_{G_j(X_j)}$. See figure (1.3) for a schematic representation of a Bell polytope and Bell inequalities.

We can now state Bell's theorem.

Bell's Theorem: No local realistic model can reproduce *all* the predictions of quantum theory.

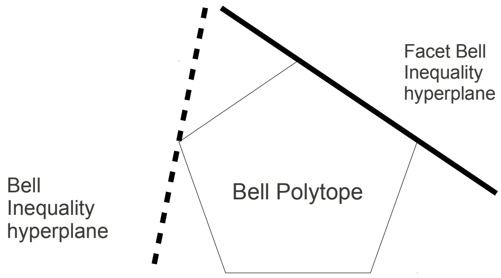

Figure 1.3: Schematic representation of the Bell polytope with hyperplanes defining Bell inequalities

To prove Bell's Theorem, therefore, we only need an example where the correlations produced by quantum theory cannot be reproduced by any local realistic model. In other words, an example where the probability vector from quantum theory lies outside the Bell Polytope $\mathcal{C}_B$. We show this explicitly for the simplest – and completely understood – Bell scenario: the CHSH scenario [6].

1.3 CHSH – the simplest Bell scenario

We will now demonstrate Bell's theorem in the simplest scenario. The Bell Polytope for this scenario is completely known and we will see an instructive instantiation of the various notions we have introduced in section (1.2). The setting we consider here is $(2, 2, 2)$, i.e., two parties ($N = 2$)– Alice and Bob – each performing two measurements ($M = 2$), where each measurement has two possible outcomes ($K = 2$). Given a combination of measurement choices of Alice and Bob, they have to produce outcomes correlated in a specified way. The quantity we are concerned with is the winning probability in this scenario, often called the CHSH game. The winning condition, of course, is that the outcomes should satisfy the CHSH correlations which we define as follows:

Let us denote Alice's measurement choice by x and Bob's measurement choice by y, where $x, y \in \{0, 1\}$. Also, we denote Alice's measurement outcome by a and Bob's measurement outcome by b, where $a, b \in \{0, 1\}$. We are interested in the conditional probabilities $P(a, b|x, y)$. There are 16 such conditional probabilities in the CHSH probability vector $\vec{P}_{(2,2,2)}$. These are subject to the following constraints:

Positivity: $P(a, b|x, y) \geq 0 \quad \forall a, b, x, y \in \{0, 1\}$

Normalization:

$$\sum_{a,b} P(a, b|x, y) = 1 \quad \forall x, y \in \{0, 1\}$$

No-signalling:

$$P(b|y) := \sum_{a} P(a, b|x, y) = \sum_{a} P(a, b|x', y) \quad \forall x \neq x', b, y \in \{0, 1\}$$

$$P(a|x) := \sum_{b} P(a, b|x, y) = \sum_{b} P(a, b|x, y') \quad \forall y \neq y', a, x \in \{0, 1\}$$

These constraints define the NS Polytope in the CHSH scenario. The winning condition in the CHSH game, i.e., the correlations that the outcomes must satisfy, are defined by:

$$a \oplus b = x.y \tag{1.10}$$

where $\oplus$ denotes addition modulo 2. Also, we define the boolean function:

$$V(a, b|x, y) = \left\{ \begin{array}{ll} 1, & \text{if } a \oplus b = x.y \\ 0, & \text{otherwise} \end{array} \right\}$$

Table (1.1) specifies the winning correlations in terms of this boolean function. From table (1.1) it is clear that the winning condition requires $a \oplus b = 0$ (i.e., agreement between a and b) for the measurement settings $(x, y) \in \{(0, 0), (0, 1), (1, 0)\}$ and $a \oplus b = 1$ (i.e., disagreement between a and b) for the measurement setting $(x, y) = (1, 1)$.

The winning probability in the CHSH game can be written as:

x	y	a	b	$V(a,b\|x,y)$
0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	0	1	1	1
0	1	0	0	1
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0

Table 1.1: Truth table for CHSH correlations

$$
\begin{aligned}
P_{win} &= \sum_{x,y,a,b} p(x)p(y)P(a,b|x,y)V(a,b|x,y) \\
&= \frac{1}{4}\sum_{x,y,a,b} P(a,b|x,y)V(a,b|x,y) \\
&= \frac{1}{4}\sum_{x,y,a,b} (P(00|00) + P(11|00) + P(00|01) + P(11|01) \\
&+ P(00|10) + P(11|10) + P(01|11) + P(10|11)) \qquad (1.11)
\end{aligned}
$$

Note that $p(x) = p(y) = \frac{1}{2} \quad \forall x, y \in \{0,1\}$, i.e., Alice and Bob choose their measurements randomly.

1.3.1 Local realistic strategy

One can visualize a local realistic strategy (i.e., a strategy that will yield correlations within the Bell Polytope) as in Fig. (1.4).

Alice's outcome for measurement $x = 0$ is a_0 and for measurement $x = 1$ it is

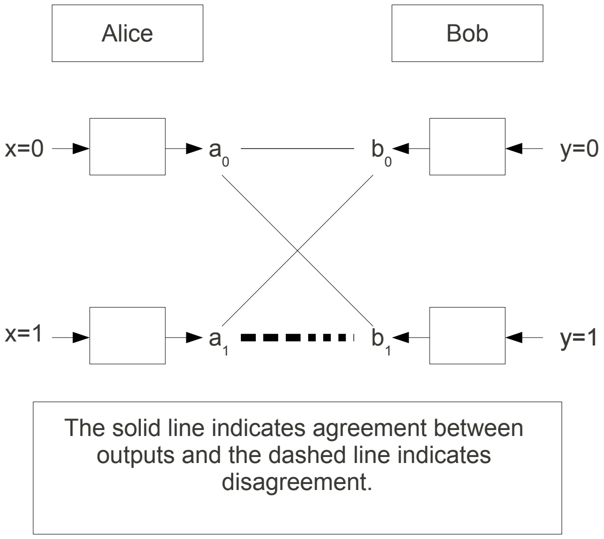

Figure 1.4: The CHSH game

a_1. Bob's outcome for measurement $y = 0$ is b_0 and for measurement $y = 1$ it is b_1. The winning condition can be written as a set of four equations:

$$\begin{aligned} a_0 \oplus b_0 &= 0, \\ a_0 \oplus b_1 &= 0, \\ a_1 \oplus b_0 &= 0, \\ a_1 \oplus b_1 &= 1. \end{aligned} \tag{1.12}$$

One can sum up any three of these equations and check that the result is inconsistent with the fourth equation. In other words, all four equations cannot be simultaneously satisfied given a pre-established local deterministic strategy – one where a_0, a_1, b_0, b_1 are assigned fixed values according to a strategy decided before the game starts. Since any probability vector within the Bell Polytope for this scenario can be written as a convex combination of these deterministic behaviours, the winning probability is bounded by $\frac{3}{4}$ or 0.75 for any local realistic (or "classical") strategy. This follows from the fact that such a strategy cannot satisfy more than 3 of the 4 winning equations (see eq. 1.12) in any single run of the CHSH game (or 'experiment'), so on average (over many repetitions of the experiment) one would win the game in 75% of the cases. That is, the classical winning probability is bounded by 75%:

$$P_{win}^{Cl} \leq 0.75 \tag{1.13}$$

We will show this explicitly in what follows.

Deterministic behaviours

Consider the single-site maps $G_i : \{0,1\} \rightarrow \{0,1\}$ that assign one of the two outcomes to each of the two measurements available to the ith party. Of course, $i \in \{1,2\}$, where $i = 1$ denotes Alice and $i = 2$ denotes Bob. So we have two single-site maps G_1 and G_2 (for Alice and Bob, respectively):

$$G_1, G_2 \in \{(00), (01), (10), (11)\} \tag{1.14}$$

where the first entry in each ordered pair (denoting a single-site map) corresponds to the outcome assigned to measurement setting 0 and the second entry to the outcome for measurement setting 1. So, for example, the single-site map (10) means that measurement 0 is mapped to outcome 1 and measurement 1 is mapped to outcome 0.

Therefore, we have 16 deterministic maps for the two parties in the CHSH scenario, i.e.,

$$\begin{aligned} G_1G_2 \in \quad & \{(00)(00), (00)(01), (00)(10), (00)(11) \\ & (01)(00), (01)(01), (01)(10), (01)(11) \\ & (10)(00), (10)(01), (10)(10), (10)(11) \\ & (11)(00), (11)(01), (11)(10), (11)(11)\} \end{aligned} \tag{1.15}$$

These deterministic maps (1.15) correspond to probability vectors $\vec{P}_{G_1G_2}$ (the extremal points of the Bell Polytope) listed in tables (1.2), (1.3), (1.4), and (1.5).

Local realistic strategy

Any local realistic strategy can be written as a convex combination of the deterministic behaviours (1.15) tabulated in (1.2), (1.3), (1.4), and (1.5).

$\vec{P}_{(2,2,2)}$	$\vec{P}_{(00)(00)}$	$\vec{P}_{(00)(01)}$	$\vec{P}_{(00)(10)}$	$\vec{P}_{(00)(11)}$
$P(00\|00)$	1	1	0	0
$P(01\|00)$	0	0	1	1
$P(10\|00)$	0	0	0	0
$P(11\|00)$	0	0	0	0
$P(00\|01)$	1	0	1	0
$P(01\|01)$	0	1	0	1
$P(10\|01)$	0	0	0	0
$P(11\|01)$	0	0	0	0
$P(00\|10)$	1	1	0	0
$P(01\|10)$	0	0	1	1
$P(10\|10)$	0	0	0	0
$P(11\|10)$	0	0	0	0
$P(00\|11)$	1	0	1	0
$P(01\|11)$	0	1	0	1
$P(10\|11)$	0	0	0	0
$P(11\|11)$	0	0	0	0
P_{win}	0.75	0.75	0.25	0.25

Table 1.2: Deterministic maps (1)

The resulting local realistic probability vector is:

$$\vec{P}^{Cl}_{(2,2,2)} = \sum_{G_1G_2} p_{G_1G_2}\vec{P}_{G_1G_2} \qquad (1.16)$$

where $p_{G_1G_2}$ is the probability weight associated with the deterministic vector $\vec{P}_{G_1G_2}$. This is represented explicitly in table (1.6).

One can now write down the classical winning probability of the game (cf. eq. 1.11 and table 1.6):

$\vec{P}_{(2,2,2)}$	$\vec{P}_{(01)(00)}$	$\vec{P}_{(01)(01)}$	$\vec{P}_{(01)(10)}$	$\vec{P}_{(01)(11)}$
$P(00\|00)$	1	1	0	0
$P(01\|00)$	0	0	1	1
$P(10\|00)$	0	0	0	0
$P(11\|00)$	0	0	0	0
$P(00\|01)$	1	0	1	0
$P(01\|01)$	0	1	0	1
$P(10\|01)$	0	0	0	0
$P(11\|01)$	0	0	0	0
$P(00\|10)$	0	0	0	0
$P(01\|10)$	0	0	0	0
$P(10\|10)$	1	1	0	0
$P(11\|10)$	0	0	1	1
$P(00\|11)$	0	0	0	0
$P(01\|11)$	0	0	0	0
$P(10\|11)$	1	0	1	0
$P(11\|11)$	0	1	0	1
P_{win}	0.75	0.25	0.75	0.25

Table 1.3: Deterministic maps (2)

$$\begin{aligned}
P_{win}^{Cl} &= \frac{3}{4}(p_{(00)(00)} + p_{(00)(01)} + p_{(01)(00)} + p_{(10)(11)} \\
&\quad + p_{(11)(10)} + p_{(01)(10)} + p_{(10)(01)} + p_{(11)(11)}) \\
&\quad + \frac{1}{4}(p_{(01)(01)} + p_{(10)(10)} + p_{(00)(10)} + p_{(11)(01)} \\
&\quad + p_{(10)(00)} + p_{(01)(11)} + p_{(00)(11)} + p_{(11)(00)}) \\
&= \frac{3}{4}(p_{(00)(00)} + p_{(00)(01)} + p_{(01)(00)} + p_{(10)(11)} \\
&\quad + p_{(11)(10)} + p_{(01)(10)} + p_{(10)(01)} + p_{(11)(11)} \\
&\quad + p_{(01)(01)} + p_{(10)(10)} + p_{(00)(10)} + p_{(11)(01)} \\
&\quad + p_{(10)(00)} + p_{(01)(11)} + p_{(00)(11)} + p_{(11)(00)}) \\
&\quad - \frac{1}{2}(p_{(01)(01)} + p_{(10)(10)} + p_{(00)(10)} + p_{(11)(01)} \\
&\quad + p_{(10)(00)} + p_{(01)(11)} + p_{(00)(11)} + p_{(11)(00)}) \\
&= \frac{3}{4} - \frac{1}{2}(p_{(01)(01)} + p_{(10)(10)} + p_{(00)(10)} + p_{(11)(01)} \\
&\quad + p_{(10)(00)} + p_{(01)(11)} + p_{(00)(11)} + p_{(11)(00)}) \\
&\leq \frac{3}{4}
\end{aligned} \tag{1.17}$$

$\vec{P}_{(2,2,2)}$	$\vec{P}_{(10)(00)}$	$\vec{P}_{(10)(01)}$	$\vec{P}_{(10)(10)}$	$\vec{P}_{(10)(11)}$
$P(00\|00)$	0	0	0	0
$P(01\|00)$	0	0	0	0
$P(10\|00)$	1	1	0	0
$P(11\|00)$	0	0	1	1
$P(00\|01)$	0	0	0	0
$P(01\|01)$	0	0	0	0
$P(10\|01)$	1	0	1	0
$P(11\|01)$	0	1	0	1
$P(00\|10)$	1	1	0	0
$P(01\|10)$	0	0	1	1
$P(10\|10)$	0	0	0	0
$P(11\|10)$	0	0	0	0
$P(00\|11)$	1	0	1	0
$P(01\|11)$	0	1	0	1
$P(10\|11)$	0	0	0	0
$P(11\|11)$	0	0	0	0
P_{win}	0.25	0.75	0.25	0.75

Table 1.4: Deterministic maps (3)

where we have used the normalization and positivity of $\vec{P}^{Cl}_{(2,2,2)}$ (cf. 1.6).

1.3.2 Quantum strategy

If, on the other hand, one was allowed to use quantum resources, then there exists a strategy that allows one to win this game with a probability higher than 0.75. That is, quantum theory can yield correlations outside the Bell Polytope of local realistic strategies. However, the maximum that quantum resources can do is to reach a winning probability close to 0.85. By "quantum resources" we refer to the subsystems that are sent to Alice and Bob on which they can perform measurements. Before the game begins, Alice and Bob can prepare the system in an optimal joint state and later (during the game) perform optimal measurements on their respective subsystems. By "optimal" we mean a choice of the joint state and local measurements that lets them do the best they can with quantum resources (which, as we will show, turns out to be a winning probability of close to 0.85).

$\vec{P}_{(2,2,2)}$	$\vec{P}_{(11)(00)}$	$\vec{P}_{(11)(01)}$	$\vec{P}_{(11)(10)}$	$\vec{P}_{(11)(11)}$
$P(00\|00)$	0	0	0	0
$P(01\|00)$	0	0	0	0
$P(10\|00)$	1	1	0	0
$P(11\|00)$	0	0	1	1
$P(00\|01)$	0	0	0	0
$P(01\|01)$	0	0	0	0
$P(10\|01)$	1	0	1	0
$P(11\|01)$	0	1	0	1
$P(00\|10)$	0	0	0	0
$P(01\|10)$	0	0	0	0
$P(10\|10)$	1	1	0	0
$P(11\|10)$	0	0	1	1
$P(00\|11)$	0	0	0	0
$P(01\|11)$	0	0	0	0
$P(10\|11)$	1	0	1	0
$P(11\|11)$	0	1	0	1
P_{win}	0.25	0.25	0.75	0.75

Table 1.5: Deterministic maps (4)

Optimal Quantum State

The optimal state for the CHSH game turns out to be the maximally entangled state of two qubits (each of which is sent to one of the two parties):

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) \tag{1.18}$$

where we choose the computational basis:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Note that $|\Psi\rangle$ lives in the tensor product Hilbert space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$, where $\mathcal{H}_\mathcal{A}$ is the Hilbert space of Alice's subsystem (qubit) and $\mathcal{H}_\mathcal{B}$ is the Hilbert space of Bob's subsystem (qubit). We have omitted the tensor product sign '$\otimes$' in $|\Psi\rangle$ but the tensor product is implied. One can now write the density matrix for the maximally entangled state:

$\vec{P}_{(2,2,2)}$	$\vec{P}^{Cl}_{(2,2,2)}$
$P(00\|00)$	$p_{(00)(00)} + p_{(00)(01)} + p_{(01)(00)} + p_{(01)(01)}$
$P(01\|00)$	$p_{(00)(10)} + p_{(00)(11)} + p_{(01)(10)} + p_{(01)(11)}$
$P(10\|00)$	$p_{(10)(00)} + p_{(10)(01)} + p_{(11)(00)} + p_{(11)(01)}$
$P(11\|00)$	$p_{(10)(10)} + p_{(10)(11)} + p_{(11)(10)} + p_{(11)(11)}$
$P(00\|01)$	$p_{(00)(00)} + p_{(00)(10)} + p_{(01)(00)} + p_{(01)(10)}$
$P(01\|01)$	$p_{(00)(01)} + p_{(00)(11)} + p_{(01)(01)} + p_{(01)(11)}$
$P(10\|01)$	$p_{(10)(00)} + p_{(10)(10)} + p_{(11)(00)} + p_{(11)(10)}$
$P(11\|01)$	$p_{(10)(01)} + p_{(10)(11)} + p_{(11)(01)} + p_{(11)(11)}$
$P(00\|10)$	$p_{(00)(00)} + p_{(00)(01)} + p_{(10)(00)} + p_{(10)(01)}$
$P(01\|10)$	$p_{(00)(10)} + p_{(00)(11)} + p_{(10)(10)} + p_{(10)(11)}$
$P(10\|10)$	$p_{(01)(00)} + p_{(01)(01)} + p_{(11)(00)} + p_{(11)(01)}$
$P(11\|10)$	$p_{(01)(10)} + p_{(01)(11)} + p_{(11)(10)} + p_{(11)(11)}$
$P(00\|11)$	$p_{(00)(00)} + p_{(00)(10)} + p_{(10)(00)} + p_{(10)(10)}$
$P(01\|11)$	$p_{(00)(01)} + p_{(00)(11)} + p_{(10)(01)} + p_{(10)(11)}$
$P(10\|11)$	$p_{(01)(00)} + p_{(01)(10)} + p_{(11)(00)} + p_{(11)(10)}$
$P(11\|11)$	$p_{(01)(01)} + p_{(01)(11)} + p_{(11)(01)} + p_{(11)(11)}$

Table 1.6: Local realistic probability vector

$$\rho = |\Psi\rangle\langle\Psi| = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \tag{1.19}$$

Optimal Quantum Measurements

Let us denote the optimal measurements that Alice and Bob perform on their subsystems by $\{A_0, A_1\}$ and $\{B_0, B_1\}$ respectively. They make spin measurements on their qubits:

$$A_0 = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad A_1 = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.20}$$

$$B_0 = \frac{\sigma_1 + \sigma_3}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad B_1 = \frac{\sigma_1 - \sigma_3}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \tag{1.21}$$

Note that the outcomes for any spin measurement are $\{+1, -1\}$, where we label $+1$ by '0' and -1 by '1'. We want to calculate the winning probability for the CHSH game given this quantum strategy:

$$P_{win}^{Q} = \frac{1}{4} \sum_{a,b,x,y \in \{0,1\}} V(a,b|x,y) P(a,b|x,y) \tag{1.22}$$

where

$$P(a,b|x,y) = \mathrm{Tr}((A_x^a \otimes B_y^b)\rho) = \langle \Psi | A_x^a \otimes B_y^b | \Psi \rangle \equiv \langle A_x^a \otimes B_y^b \rangle$$

$$A_x^a = \frac{\mathbb{I} + (-1)^a A_x}{2} \tag{1.23}$$

$$B_y^b = \frac{\mathbb{I} + (-1)^b B_y}{2} \tag{1.24}$$

Clearly,

$$\langle A_x^a \otimes B_y^b \rangle = \frac{1}{4} \langle \mathbb{I} \otimes \mathbb{I} + (-1)^b \mathbb{I} \otimes B_y + (-1)^a A_x \otimes \mathbb{I} + (-1)^{a \oplus b} A_x \otimes B_y \rangle \tag{1.25}$$

Let $a \oplus b = c$. Then $b = a \oplus c$. We shall use it in the following:

$$\begin{aligned}
P_{win}^{Q} &= \frac{1}{4} \sum_{a,b,x,y \in \{0,1\}} V(a,b|x,y) \langle A_x^a \otimes B_y^b \rangle & (1.26)\\
&= \frac{1}{4} \sum_{a,c,x,y \in \{0,1\}} V(c|x,y) \langle A_x^a \otimes B_y^b \rangle & (1.27)\\
&= \frac{1}{4} \sum_{a,c,x,y \in \{0,1\}} V(c|x,y) \frac{1}{4} \langle \mathbb{I} \otimes \mathbb{I} + (-1)^{a \oplus c} \mathbb{I} \otimes B_y + (-1)^a A_x \otimes \mathbb{I} & \\
&\qquad + (-1)^c A_x \otimes B_y \rangle & (1.28)\\
&= \frac{1}{2} (1 + \frac{\langle A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1 \rangle}{4} & (1.29)\\
&= \frac{1}{2} (1 + \frac{\langle CHSH \rangle}{4}) & (1.30)
\end{aligned}$$

where

$$\langle CHSH \rangle := \langle A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1 \rangle.$$

Consider spin measurement on Alice's qubit along $\hat{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ axis (i.e., measurement of $A^{\alpha} = \vec{\sigma}.\hat{\alpha}$), and measurement on Bob's qubit along $\hat{\beta} =$

$(\beta_1, \beta_2, \beta_3)$ axis (i.e., measurement of $B^\beta = \vec{\sigma}.\hat{\beta}$). Of course, $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, where σ_1, σ_2, and σ_3 are the three Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.31}$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{1.32}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.33}$$

corresponding to spin measurements along the X, Y, and Z axis respectively.

Now,

$$\langle A^\alpha \otimes B^\beta \rangle = \mathrm{Tr}((A^\alpha \otimes B^\beta)\rho) = \alpha_1\beta_1 - \alpha_2\beta_2 + \alpha_3\beta_3 \tag{1.34}$$

Denoting any spin measurement $\vec{\sigma}.\hat{\gamma}$ by the corresponding unit vector $(\gamma_1, \gamma_2, \gamma_3)$ along which the measurement is made, we have

$$A_0 = \sigma_1 \to (1, 0, 0), \quad A_1 = \sigma_3 \to (0, 0, 1) \tag{1.35}$$

$$B_0 = \frac{\sigma_1 + \sigma_3}{\sqrt{2}} \to (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}), \quad B_1 = \frac{\sigma_1 - \sigma_3}{\sqrt{2}} = (\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}) \tag{1.36}$$

$$\begin{aligned} \langle CHSH \rangle &= \langle A_0 \otimes B_0 \rangle + \langle A_0 \otimes B_1 \rangle + \langle A_1 \otimes B_0 \rangle - \langle A_1 \otimes B_1 \rangle \\ &= (1,0,0).(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) + (1,0,0).(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}) \\ &\quad +(0,0,1).(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}) - (0,0,1).(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}) \\ &= 2\sqrt{2} \end{aligned} \tag{1.37}$$

$$\therefore P^Q_{win} = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.85 \tag{1.38}$$

1.3.3 Tsirelson's bound

It turns out that $P^Q_{win} = \frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 0.85$ (cf. 1.38) is the maximum winning probability in the CHSH game if quantum resources are allowed. In terms of the CHSH operator:

$$\langle CHSH\rangle \leq 2\sqrt{2} \tag{1.39}$$

This fact was shown by Tsirelson [7] and is known as Tsirelson's bound. We will prove this inequality. Note that classically (or local realistically),

$$\langle CHSH\rangle \leq 2 \tag{1.40}$$

This is the Bell-CHSH inequality. It follows from the fact that no pre-decided fixed assignment to A_0, A_1, B_0, B_1 of values in $[-1, 1]$ can make $\langle CHSH\rangle$ exceed 2. This is an instance of Bell's theorem – quantum theory can violate this Bell Inequality up to a maximum value of $2\sqrt{2}$.

Proof.

Consider four arbitrary spin measurement operators:

$$A_0 = \vec{a}_0.\vec{\sigma}, A_1 = \vec{a}_1.\vec{\sigma}, B_0 = \vec{b}_0.\vec{\sigma}, B_1 = \vec{b}_1.\vec{\sigma}$$

and the CHSH operator

$$\langle CHSH\rangle = \langle A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1\rangle.$$

Now,

$$\begin{aligned}
& (A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1)^2 && (1.41)\\
= \;& 4\mathbb{I} \otimes \mathbb{I} + A_1A_0 \otimes B_0B_1 - A_0A_1 \otimes B_0B_1 \\
& +A_0A_1 \otimes B_1B_0 - A_1A_0 \otimes B_1B_0 && (1.42)\\
= \;& 4\mathbb{I} \otimes \mathbb{I} + [A_1, A_0] \otimes B_0B_1 + [A_0, A_1] \otimes B_1B_0 && (1.43)\\
= \;& 4\mathbb{I} \otimes \mathbb{I} + [A_1, A_0] \otimes [B_0, B_1] && (1.44)
\end{aligned}$$

Using the fact that

$$\langle \hat{O}\rangle \leq \sqrt{\langle \hat{O}^2\rangle} \tag{1.45}$$

where $\hat{O}$ is a Hermitian operator on $\mathcal{H}_A \otimes \mathcal{H}_B$, we have

$$\begin{aligned}
\langle CHSH\rangle &\leq \sqrt{\langle 4\mathbb{I} \otimes \mathbb{I} + [A_1, A_0] \otimes [B_0, B_1]\rangle} && (1.46)\\
&\leq \sqrt{4 + 4} \\
&= 2\sqrt{2} && (1.47)
\end{aligned}$$

Note that if the local observables are commuting, i.e., $[A_1, A_0] = 0$ or $[B_0, B_1] = 0$, then we recover the classical bound $\langle CHSH \rangle \leq 2$. Clearly, the noncommutativity of observables in quantum mechanics plays a crucial role in the quantum advantage, yielding a winning probability of 85% compared to the classical 75% in the CHSH game.

1.4 The General Problem

The task, in a nutshell, is to characterize the following classes of correlations in the Bell scenario:

Local: This comprises the set of correlations that can be reproduced by a local hidden variable model. This set is defined by the existence of a decomposition (1.1):

$$\begin{aligned} & P(A_1, A_2, ..., A_N | X_1, X_2, ..., X_N) \\ = & \sum_{\lambda} P(\lambda) P(A_1 | X_1, \lambda) P(A_2 | X_2, \lambda), ..., P(A_N | X_N, \lambda) \end{aligned} \tag{1.48}$$

These correlations form the Bell Polytope. Operationally, local correlations are those that can be established via pre-shared randomness (λ).

Quantum: This comprises the set of correlations that can be reproduced by quantum theory. In general terms, the quantum set of correlations is defined by the existence of a valid N-partite quantum state ρ (shared by the N parties in the Bell scenario) and measurement operators $M_{X_i}^{A_i}$ such that:

$$\begin{aligned} & P(A_1, A_2, ..., A_N | X_1, X_2, ..., X_N) \\ = & \operatorname{Tr}((M_{X_1}^{A_1} \otimes M_{X_2}^{A_2} \otimes ... \otimes M_{X_N}^{A_N}) \rho) \end{aligned} \tag{1.49}$$

So far, the operational meaning of quantum correlations is not completely clear. While local correlations and no-signalling correlations are specified by explicit constraints arising from physical restrictions – local realism in the former case, and no-signalling (from special relativity) in the latter, such an operational meaning is not very clear in the case of quantum correlations. Indeed, recent efforts to obtain an information-theoretic operational meaning include information causality [8] and non-trivial communication complexity [9], and we will review some of these recent attempts in the next chapter.

No-signalling: This comprises the set of correlations that obey the three minimal conditions of *Positivity*, *Normalization*, and *No-signalling* (see 1.2.2). This set of correlations forms the **Non-signalling (NS) Polytope**. Operationally, this just means that the measurement statistics for any subset of parties in the Bell scenario does not depend on the measurements made by the remaining parties.

The set of local correlations is strictly within the set of quantum correlations which in turn is strictly within the set of no-signalling correlations. The general problem is to find the boundaries of these sets of correlations, especially the quantum set, for *any* Bell scenario. This question – characterizing these correlations – will be taken up in the next chapter.

1.5 Chapter Summary

We have introduced the notion of Bell polytopes, Bell inequalities, and nonlocal correlations. Our discussion has been limited to the basic concepts required to understand nonlocality. For a fairly exhaustive treatment of Bell Inequalities, including some of recent advances in the field, one may refer to [5] and [10] which I perused in my survey of the literature. See [11] for a list of open questions in the field. The formal problem statement, “All the Bell Inequalities”, can be found on the Hannover page [12].

Chapter 2

Quantum nonlocality from physical principles?

2.1 Introduction

That nature allows nonlocal correlations (contradicting local realism) is an experimental fact [13] and any theory that seeks to describe nature will have to allow for nonlocality. In this sense, then, nonlocality is a theory-independent notion. It refers to properties of correlations that can be observed in some well-defined correlation experiments (often called 'Bell tests') that we discussed in the last chapter. The theory which models these experiments predicts the extent to which Bell inequalities may be violated. Quantum theory allows a certain degree of nonlocality which is not maximal. In this chapter we will review recent results that consider the possibility of Bell inequality violations beyond quantum theory, especially the so-called PR-boxes, in an attempt to answer the question: Why is quantum theory not maximally nonlocal? After all, all no-signalling correlations are consistent with special relativity, yet only some of them are allowed in quantum theory. What rules out the other no-signalling correlations? Answering this question would involve identifying physical principle(s) underlying quantum theory that characterize quantum correlations.

Should a restriction on nonlocality be considered a fundamental axiom in any physical theory [14]? Or, is it possible to explain such a restriction as arising from some more fundamental feature(s) of the physical world? Various authors have proposed physical principles that might help in explaining the restrictions on quantum correlations. One of the motivations of this research endeavour is to find physical principle(s) that will single out quantum theory in a class of conceivable nonlocal theories that may be similar to quantum theory in other respects.

The approaches we will review are Information Causality[8], Macroscopic Locality [18] and fine-grained uncertainty relations [19].

2.2 Why Tsirelson's bound?

Algebraically, the maximum allowed value of $\langle CHSH \rangle$ is 4. Equivalently, the maximum winning probability in the CHSH game can be 1. However, we know from Tsirelson's bound that quantum theory cannot give us these maximal correlations. A natural question that arises is what law of nature might forbid such correlations. Popescu and Rohrlich [14] asked this question and came up with a joint probability distribution for the CHSH scenario, now known as the PR-box [20] in the literature, which is consistent with no-signalling constraints and gives a winning probability of 1 in the CHSH game, or equivalently, a value of 4 for $\langle CHSH \rangle$. The PR-box for the CHSH scenario is a hypothetical NS-box producing the following correlations:

$$P(a, b|x, y) = \frac{1}{2}\delta_{a\oplus b,xy} \tag{2.1}$$

where $\delta_{a\oplus b,xy} = 1$, if $a \oplus b = xy$, 0 otherwise. We list the probability vector explicitly in table (2.1). Clearly, this is a no-signalling probability distribution, with $P(a|x) = P(b|y) = \frac{1}{2} \quad \forall a, b, x, y \in \{0, 1\}$, that gives a winning probability of 1 in the CHSH game. In spite of being non-signalling, such a correlation is not accessible in quantum theory.

Several authors have studied the consequences of allowing no-signalling correlations beyond quantum theory, and we briefly mention some of these before turning to the three recent results we will mainly review – information causality, macroscopic locality, and fine-grained uncertainty relations. We list some results about PR-boxes (see [20] for an overview):

1. **Simulation of the singlet:** Correlations of the singlet (maximally entangled state of two qubits) can be simulated by supplementing local hidden variables with *a single use of the PR-box* [21]. Since no communication is required, this result was an improvement over an earlier result of Toner and Bacon [22] showing that correlations of the singlet can be simulated by local hidden variables supplemented with *a single bit of communication.* Note that the PR-box is non-signalling and

$\vec{P}_{PR}$	$P(a,b\|x,y)$
$P(00\|00)$	1/2
$P(01\|00)$	0
$P(10\|00)$	0
$P(11\|00)$	1/2
$P(00\|01)$	1/2
$P(01\|01)$	0
$P(10\|01)$	0
$P(11\|01)$	1/2
$P(00\|10)$	1/2
$P(01\|10)$	0
$P(10\|10)$	0
$P(11\|10)$	1/2
$P(00\|11)$	0
$P(01\|11)$	1/2
$P(10\|11)$	1/2
$P(11\|11)$	0

Table 2.1: PR-box probability vector

is therefore a strictly weaker resource than one bit of communication (which requires signalling).

2. **Communication Complexity:** Wim van Dam [9] demonstrated that allowing a PR-box as a resource makes the communication complexity of a task trivial. To be more precise, Alice receives an n-bit string $\vec{x}$ and Bob receives n-bit string $\vec{y}$ and they are supposed to evaluate the Boolean function $f(\vec{x},\vec{y})$. In the particular case $f(\vec{x},\vec{y}) = \sum_{i=0}^{n-1} x_i y_i$, it is known that Alice must communicate all her n bits to Bob for him to produce the correct output. Even shared entanglement doesn't offer an advantage. However, given a PR-box, Alice and Bob only need to sequentially input their respective bits (x_i and y_i, $\quad \forall i = 0, 1, ..., n-1$) into the PR-box to produce correlated outputs a_i and b_i satisfying $a_i \oplus b_i = x_i y_i$. All they need to evaluate is $\sum_i (a_i \oplus b_i)$. Bob can calculate $\sum_i b_i$ and Alice needs to communicate *one bit* $\sum_i a_i$ to Bob for him to evalute the given function. Thus the communication complexity of the task is reduced from n bits to one bit, i.e., it becomes trivial. See [9] for details.

3. **Trivial reversible dynamics:** In [24], David Gross and co-authors showed that in a theory allowing maximally nonlocal correlations (i.e.,

the PR-box type correlations), if one requires reversibility of dynamics, then the allowed transformations in a such a theory would be highly restricted (indeed, "trivial") compared to quantum theory – it can be efficiently simulated classically. They prove this result in the framework of generalized probabilistic theories [23].

2.3 Information Causality

In this section we will review a proposal that appeared in 2009 [8], formulating a physical principle called 'information causality' that quantum correlations are shown to obey, and which is violated by stronger correlations. It offers a possible way to distinguish the set of quantum correlations from stronger non-signalling correlations. This principle, however, turns out to have some limitations [30] that are in fact generic to any bipartite physical principle (like no-signalling, of which information causality is a generalization).

2.3.1 Information Causality as a physical principle

The basic setting of information causality is in terms of a game between two players – Alice and Bob. Alice gets a string of N bits $\vec{a} = \{a_0, a_1, ..., a_{N-1}\}$, where $Pr(a_i = 0) = Pr(a_i = 1) = \frac{1}{2}, \quad \forall i \in \{0, 1, ..., N-1\}$, i.e., Alice gets N random and independent bits as input. The task that Alice and Bob have to accomplish is the following:

Task: Given a fixed amount of classical communication from Alice to Bob (m bits, where $0 \leq m \leq N$), Alice wants to grant Bob access to as many bits of her N-bit string, $\vec{a}$, as possible. They are allowed to share a no-signalling resource of their choice before the game starts, i.e., before Alice receives the input $\vec{a}$.

One can think of the pre-shared no-signalling resource as a pair of correlated boxes, one with Alice and the other with Bob. The correlations between them should satisfy positivity, normalization, and no-signalling. Before the game starts, Alice and Bob agree on some protocol according to which they will use their local resources during the game (when communication is limited to m bits from Alice to Bob). Once Alice receives the string $\vec{a}$, she makes an appropriate input (according to the pre-decided protocol) to her no-signalling box ("NS-box") and processes the output it produces to obtain an m-bit string $\vec{x} = \{x_0, x_1, ..., x_{m-1}\}$. She sends these m classical bits to Bob. Bob, on the other hand, receives a random input $b \in \{0, 1, ..., N-1\}$, i.e., $Pr(b) = \frac{1}{N}, \quad \forall \quad b \in \{0, 1, ..., N-1\}$. Bob is supposed to retrieve the bit a_b from Alice's input string $\vec{a}$. The resources he has for this purpose include his NS-box and the classical communication $\vec{x}$ from Alice. He makes an appropriate input to his NS-box and processes the output according to the pre-decided protocol to give an answer β for the value of a_b. If $\beta = a_b$, Bob has successfully retrieved the required bit in $\vec{a}$. We will quantify the efficiency of this task in terms of mutual information between the bits of $\vec{a}$ and Bob's guess β for them. See Fig. (2.1) for a schematic representation of the Information Causality (IC) game.

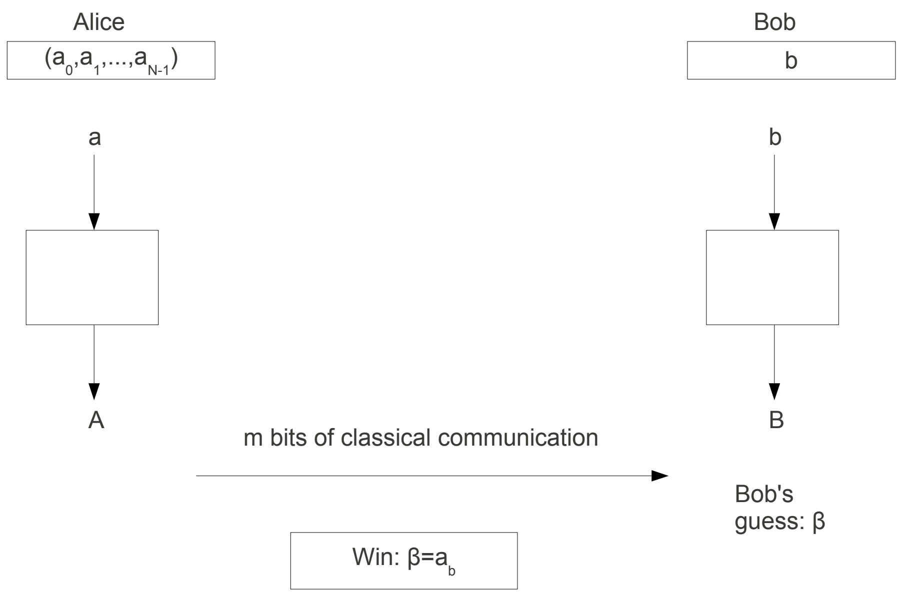

Figure 2.1: The IC game

Efficiency: The efficiency of any protocol according to which Alice and Bob play the game is quantified by Bob's **accessible information**:

$$I \equiv \sum_{K=0}^{N-1} I(a_K : \beta | b = K) \tag{2.2}$$

$I(a_K : \beta|b = K)$ is the mutual information between Bob's guess β, given $b = K$, and the value of the bit a_b. Note that this accessible information is calculated from *purely classical data* – the inputs and outputs from the NS-boxes – and does not depend on the internal working of the NS-boxes. Indeed, this "blackbox" approach lets us make theory-independent statements, relating only information-theoretic quantities. Information Causality is such a statement.

On the other hand, the mutual information $I(\vec{a} : \vec{x}, B)$ between Alice's string $\vec{a}$ and everything Bob possesses – the string $\vec{x}$ communicated by Alice and B, his part of the pre-shared correlation – is a quantity that depends on the underlying theory describing the NS-boxes, and it is not clear that such a mutual information can always be defined for any arbitrary no-signalling theory. Instead of worrying about the structure of the no-signalling theory governing our NS-boxes, we focus on whether a mutual information satisfying the following three basic properties can be defined:

1. **Consistency**: If subsystems A and B represent classical random variables, then $I(A : B)$ reduces to the standard Shannon mutual information.

2. **Data processing inequality**: This requires that any local data processing cannot increase the mutual information between two parties, i.e., if A and B are the two subsystems and $B \to B'$ is a transformation allowed in the theory, then after this local transformation, $I(A : B) \geq I(A : B')$.

3. **Chain rule**: This requires that a conditional mutual information, $I(A : B|C)$, can be defined, obeying the following identity for all A, B, C: $I(A : B, C) = I(A : B|C) + I(A : C)$. This implies the following identity: $I(A : B, C) - I(A : C) = I(A : B|C) = I(A, C : B) - I(B : C)$.

We state the principle of information causality as follows:

Principle of Information Causality: If a mutual information obeying the three elementary properties above can be defined, then: (1) Information Causality holds, i.e., $I(\vec{a} : \vec{x}, B) \leq m$, and (2) $I(\vec{a} : \vec{x}, B) \geq I$.

Therefore, a **necessary condition** for Information Causality is: $I \leq m$.

It turns out that this principle is satisfied by classical and quantum correlations, and violated by no-signalling correlations stronger than quantum correlations. Indeed, this principle even lets us derive Tsirelson's bound [7] in an information-theoretic manner, without appealing to the Hilbert space structure of quantum theory. We state the proof for information causality, and related corollaries, given in [8]:

Proof.

1) $I(\vec{a}:\vec{x},B) \geq I$

$$\begin{aligned}
I(a_0,a_1,...,a_{N-1}:\vec{x},B) &= I(a_0:\vec{x},B)+I(a_1,...,a_{N-1}:\vec{x},B|a_0) \\
&= I(a_0:\vec{x},B)+I(a_1,...,a_{N-1}:\vec{x},B,a_0)-I(a_1,...,a_{N-1}:a_0) \\
&= I(a_0:\vec{x},B)+I(a_1,...,a_{N-1}:\vec{x},B,a_0) \\
&\quad \text{(using the mutual independence of bits in } \vec{a}) \\
&\geq I(a_0:\vec{x},B)+I(a_1,...,a_{N-1}:\vec{x},B) \\
&\quad \text{(using the Data Processing Inequality)}
\end{aligned}$$

Similarly iterating this process $N-1$ times, we get

$$\begin{aligned}
I(a_1,...,a_{N-1}:\vec{x},B) &\geq I(a_1:\vec{x},B)+I(a_2,...,a_{N-1}:\vec{x},B) \\
&\;\;. \\
&\;\;. \\
&\;\;. \\
I(a_{N-2},a_{N-1}:\vec{x},B) &\geq I(a_{N-2}:\vec{x},B)+I(a_{N-1}:\vec{x},B)
\end{aligned}$$

$$\therefore \quad I(\vec{a}:\vec{x},B) \geq \sum_{K=0}^{N-1} I(a_K:\vec{x},B)$$

Now, $I(a_K:\vec{x},B) \geq I(a_K:\beta|b=K)$ (using data processing inequality, since Bob processes $\vec{x}$ from Alice, and his part B of the no-signalling box, given $b=K$, to compute β), and therefore

$$I(\vec{a}:\vec{x},B) \geq \sum_{K=0}^{N-1} I(a_K:\beta|b=K) \equiv I \tag{2.3}$$

2) $I(\vec{a}:\vec{x},B) \leq m$: Firstly, from consistency and the data processing inequality, the mutual information $I(A:B)$ satisfies an important property of Shannon mutual information: $I(A:B)=0$ if systems A and B are independent. This follows from considering the fact that, beginning with two classical independent systems (with $I(A' : B') = 0$ by consistency), one can process them locally ($A' \to A, B' \to B$) to produce independent nonclassical systems. By the data processing inequality, we have $I(A : B) \leq 0$. Also, $I(A : B) \geq 0$ by the non-negativity of mutual information. Therefore, $I(A : B) = 0$ for independent systems A and B. Now,

$$\begin{aligned} I(\vec{a} : \vec{x}, B) &= I(\vec{a} : B) + I(\vec{a} : \vec{x}|B) \\ &= I(\vec{a} : \vec{x}|B) \text{ } (\vec{a} \text{ and } B \text{ are independent}) \\ &= I(\vec{x} : \vec{a}, B) - I(\vec{x}, B) \\ &\leq I(\vec{x} : \vec{a}, B) \text{ (because } I(\vec{x}, B) \geq 0) \\ &\leq I(\vec{x} : \vec{x}) \end{aligned}$$

where the last inequality follows from data processing: the joint state of $\vec{x}, \vec{a}, B$ is a convex combination of joint states of $\vec{a}$ and B for every value $\vec{x}$ takes according to some distribution. Hence, $\vec{x} \to \vec{a}, B$ is an allowed transformation. Clearly,

$$I(\vec{x} : \vec{a}, B) \leq I(\vec{x} : \vec{x}) \leq m$$

Thus,

$$I(\vec{a} : \vec{x}, B) \leq m \tag{2.4}$$

3) Lower bound on I

$$I(a_K : \beta | b = K) = h(a_K) - H(a_K|\beta, b = K) \tag{2.5}$$

Now,

$$h(a_K) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$$

(since $Pr(a_K = 0) = Pr(a_K = 1) = \frac{1}{2}$. The bits in the string $\vec{a}$ are independent and random.)

$$H(a_K|\beta, b = K) = H(a_K \oplus \beta|\beta, b = K) \leq H(a_K \oplus \beta|b = K) = h(P_K)$$

where

$$P_K = Pr(a_K \oplus \beta = 0|b = K)$$

is the probability that Bob guesses the required bit a_b correctly. Thus,

$$\begin{aligned} I &= \sum_{K=0}^{N-1} I(a_K : \beta | b = K) && (2.6) \\ &= \sum_{K=0}^{N-1} (1 - H(a_K|\beta, b = K)) && (2.7) \\ &= N - \sum_{K=0}^{N-1} H(a_K|\beta, b = K) && (2.8) \\ &\geq N - \sum_{K=0}^{N-1} h(P_K) && (2.9) \end{aligned}$$

Using eq. (2.4), we have $N - \sum_{K=0}^{N-1} h(P_K) \leq I \leq m$. Therefore, information causality limits the probability that Bob correctly retrieves the required bit:

$$\sum_{K=0}^{N-1} h(P_K) \geq N - m \qquad (2.10)$$

4) Information Causality holds in classical and quantum physics: To verify that the principle of information causality is obeyed by classical and quantum physics, we need to show that the mutual information and conditional mutual information in quantum physics satisfy the three elementary properties required of them for information causality to hold. Since classical correlations form a subset of quantum correlations, we need to verify information causality only for the quantum case. In quantum theory, the von Neumann entropy $S(\rho)$ of a state ρ is defined as:

$$S(\rho) = -\mathrm{Tr}\rho \log \rho$$

This reduces to the classical Shannon entropy in the eigenbasis of ρ. For any tripartite state ρ_{ABC} with reduced states $\rho_A = \mathrm{Tr}_{BC}\rho_{ABC}$, $\rho_B = \mathrm{Tr}_{AC}\rho_{ABC}$, and $\rho_C = \mathrm{Tr}_{AB}\rho_{ABC}$,

$$\begin{aligned} I(A : B) &= S(\rho_A) + S(\rho_B) - S(\rho_{AB}) && (2.11) \\ I(A : B|C) &= S(\rho_{AC}) + S(\rho_{BC}) - S(\rho_{ABC}) - S(\rho_C) && (2.12) \end{aligned}$$

These expressions are invariant with respect to exchanging A and B. From the strong subadditivity of quantum entropy [15], we know:

$$S(\rho_{ABC}) + S(\rho_C) \leq S(\rho_{AC}) + S(\rho_{BC}) \tag{2.13}$$

If subsystem C is in a pure state ρ_C (so that $S(\rho_C) = 0$), this reduces to

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B) \tag{2.14}$$

Clearly, the mutual information and the conditional mutual information in quantum theory are non-negative due to strong subadditivity. The conditions of consistency, data processing inequality, and the chain rule hold for mutual information in quantum theory. Hence, quantum theory satisfies information causality.

5) Violation of Information Causality: We demonstrate a protocol using PR-boxes – no-signalling resources not accessible within quantum theory – that violates information causality, making the entire string $\vec{a}$ received by Alice accessible to Bob with only one bit of communication ($m = 1$). Of course, the amount Bob can extract is limited to one bit, but he can extract *any* one bit of his choice – the whole string $\vec{a}$ is open for a readout and his accessible information is therefore N bits.

Consider the simplest scenario: Alice receives a string of two bits $\vec{a} = \{a_0, a_1\}$ and Bob receives a single bit $b \in \{0, 1\}$. The task is to retrieve a_b. Now, we allow Alice and Bob the use of a PR-box, producing the correlations $A \oplus B = ab$ for inputs $a, b \in \{0, 1\}$ (respectively) and outputs $A, B \in \{0, 1\}$ (respectively) for Alice and Bob. Their strategy is the following:

1. When Alice receives $\vec{a}$, she inputs $a = a_0 \oplus a_1$ to her part of the PR-box, obtains output A, and sends a one bit message $x = a_0 \oplus A$ to Bob.
2. Bob inputs the bit b he receives into his PR-box and obtains a one-bit output B.
3. Given his output B and the message x from Alice, Bob computes his guess for a_b: $\beta = x \oplus B = a_0 \oplus A \oplus B$.

The probability that Bob guesses a_0 correctly is given by

$$P_I = \frac{1}{2}[P(A \oplus B = 0|a = 0, b = 0) + P(A \oplus B = 0|a = 1, b = 0)]$$

and the probability of guessing a_1 correctly is

$$P_{II} = \frac{1}{2}[P(A \oplus B = 0|a = 0, b = 1) + P(A \oplus B = 1|a = 1, b = 1)]$$

($\beta = a_0 \oplus A \oplus B$, so one needs $A \oplus B = 0$ to retrieve a_0 when $b = 0, a \in \{0, 1\}$. Also, one needs $A \oplus B = 0$ when $b = 1, a = 0$, and $A \oplus B = 1$ when $b = 1, a = 1$.)

This defines the following CHSH parameter:

$$S = \sum_{a=0}^{1} \sum_{b=0}^{1} P(A \oplus B = ab|a, b) = 2(P_I + P_{II}) \tag{2.15}$$

This is related to the winning probability in a CHSH game (cf. 1.11) according to:

$$P_{win} = \frac{1}{4} S.$$

So, classically $S \leq S_C = 3$, and for quantum correlations $S \leq S_Q = 2 + \sqrt{2}$ (Tsirelson's bound). The maximum algebraic value, $S = 4$, is achieved by a PR-box. Since the box Alice and Bob share is a PR-box, we can compute Bob's guess:

$$\begin{aligned} \beta &= x \oplus B \\ &= a_0 \oplus A \oplus B \\ &= a_0 \oplus ab \\ &= a_0 \oplus (a_0 \oplus a_1) b \end{aligned} \tag{2.16}$$

Clearly,

$$\beta = \left\{ \begin{array}{ll} a_0, & \text{if } b = 0 \\ a_1, & \text{if } b = 1 \end{array} \right\}$$

See figure (2.2).

So Bob can retrieve any bit from Alice's input perfectly, his accessible information being $I = 2$, clearly violating a necessary condition for information causality, $I \leq 1$. Information causality, thus, rules out PR-boxes. We will now consider a more general set of isotropic correlations (which are nonsignalling) that yield the PR-box as a special case. Indeed, any NS-box can be brought into an isotropic form (keeping S fixed) [17] (cf. table 2.2), satisfying:

$$P(A \oplus B = ab|a, b) = \frac{1}{2}(1 + E), \quad \text{where } 0 \leq E \leq 1 \tag{2.17}$$

For $E = 1$, we have the PR-box, which is maximally nonlocal. $E = 0$ corresponds to completely random (therefore uncorrelated) outputs. This NS-box violates the classical bound $S \leq 3$ as soon as $E > \frac{1}{2}$. Tsirelson's

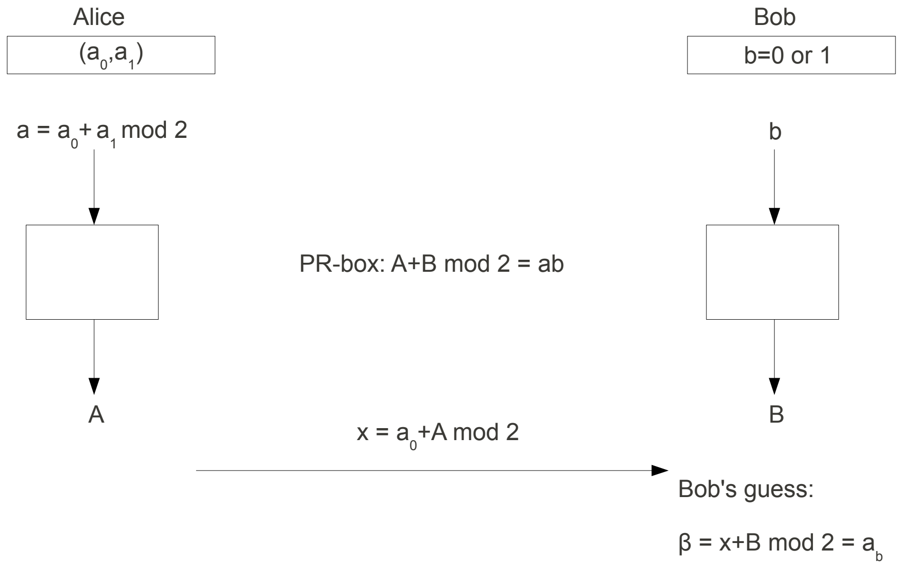

Figure 2.2: The IC game using a PR-box

bound is equivalent to $E \leq \frac{1}{\sqrt{2}}$. We will show that information causality is violated as soon as Tsirelson's bound is exceeded, i.e., for $E > \frac{1}{\sqrt{2}}$.

Consider a scenario where Alice receives a string $\vec{a} = (a_0, ..., a_{N-1})$ of $N = 2^n$ bits, where n is an integer. Bob receives an n-bit index string $(b_0, b_1, ..., b_{n-1})$ where $b = \sum_{k=0}^{n-1} b_k 2^k$ is the index of the bit to be retrieved from $\vec{a}$. One bit of communication is allowed from Alice to Bob, i.e., $m = 1$. The function that Alice and Bob are required to compute is $f_n(\vec{a}, b) \equiv a_b$. We know that for $n = 1$, this function is given by

$$f_1((a_0, a_1), b_0) = a_0 \oplus (a_0 \oplus a_1)b_0$$

Computing this function required one PR-box. For $n > 1$, $N = 2^n$, we split the N-bit string $\vec{a}$ into $\frac{N}{2}$-bit strings $\vec{a}'$ and $\vec{a}''$: $\vec{a} = \vec{a}'\vec{a}''$. The function $f_n(\vec{a}, b)$ can now be recursively obtained:

$$f_n(\vec{a}, b) = f_{n-1}(\vec{a}', b') \oplus b_{n-1}[f_{n-1}(\vec{a}', b') \oplus f_{n-1}(\vec{a}'', b')] \tag{2.18}$$

$\vec{P}_{iso}$	$P_{iso}(A,B\|a,b)$
$P(00\|00)$	$\frac{1}{4}(1+E)$
$P(01\|00)$	$\frac{1}{4}(1-E)$
$P(10\|00)$	$\frac{1}{4}(1-E)$
$P(11\|00)$	$\frac{1}{4}(1+E)$
$P(00\|01)$	$\frac{1}{4}(1+E)$
$P(01\|01)$	$\frac{1}{4}(1-E)$
$P(10\|01)$	$\frac{1}{4}(1-E)$
$P(11\|01)$	$\frac{1}{4}(1+E)$
$P(00\|10)$	$\frac{1}{4}(1+E)$
$P(01\|10)$	$\frac{1}{4}(1-E)$
$P(10\|10)$	$\frac{1}{4}(1-E)$
$P(11\|10)$	$\frac{1}{4}(1+E)$
$P(00\|11)$	$\frac{1}{4}(1-E)$
$P(01\|11)$	$\frac{1}{4}(1+E)$
$P(10\|11)$	$\frac{1}{4}(1+E)$
$P(11\|11)$	$\frac{1}{4}(1-E)$

Table 2.2: Isotropic distribution

where $b' \equiv b_{n-2}b_{n-1}...b_0$, $b \equiv b_{n-1}b_{n-2}...b_0$. For the case $n = 2$, we need to use three PR-boxes:

$$\begin{aligned} & f_2((a_0, a_1, a_2, a_3), (b_0, b_1)) \\ &= f_1((a_0, a_1), b_0) \oplus b_1[f_1((a_0, a_1), b_0) \oplus f_1((a_2, a_3), b_0)] \\ &= a_0 \oplus (a_0 \oplus a_1)b_0 \oplus b_1[a_0 \oplus (a_0 \oplus a_1)b_0 \oplus a_2 \oplus (a_2 \oplus a_3)b_0] \end{aligned}$$

We label the three PR-boxes as Box_1,Box_2, and Box_3. To accomplish this task (for $n = 2$) Alice and Bob use the following protocol:

1. Alice inputs $a_0 \oplus a_1$ to her part of Box_1 and gets output A_1, $a_2 \oplus a_3$ to her part of Box_2 to get output A_2, and $(A_1 \oplus a_0 \oplus a_2 \oplus A_2)$ to her part of Box_3 to get output A_3.

2. If $b_1 = 0$, Bob inputs b_0 to his part of Box_1 to get output B_1. Otherwise $(b_1 = 1)$, he inputs b_0 to his part of Box_2 to get B_2. Bob inputs b_1 to his part of Box_3 to get B_3.

3. Alice communicates one bit $x = a_0 \oplus A_1 \oplus A_3$ to Bob.

4. Bob computes his guess for a_b: $\beta = x \oplus B_1\delta_{b_1,0} \oplus B_2\delta_{b_1,1} \oplus B_3$.

Using the PR-box correlation, one can easily verify that $\beta = a_b$ for all $b \equiv b_1 b_0 \in \{00, 01, 10, 11\}$. This is equivalent to computing the function f_2, which may also be verified by substituting values of b. See figure (2.3).

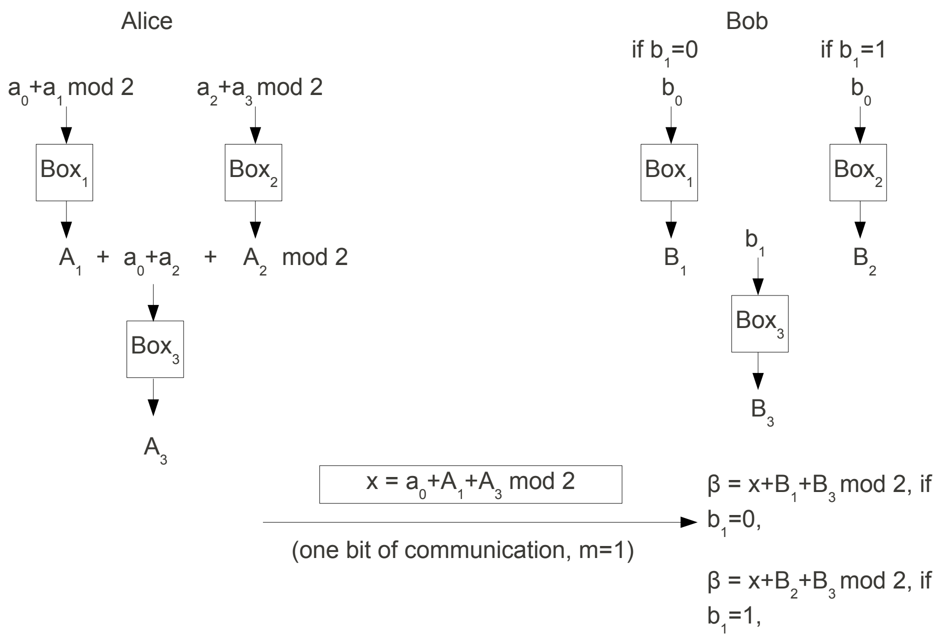

Figure 2.3: The IC game with three PR-boxes

In general, Alice and Bob require $N - 1$ (where $N = 2^n$) PR-boxes to compute f_n by appropriately nesting them (see [8] for details of the nesting protocol and figure (2.4) for a schematic representation).

Now we consider replacing the PR-boxes by isotropic no-signalling boxes (cf. 2.2). In the $n = 1$ case, the probability of guessing a_0 and a_1 correctly is given (respectively) by $P_I = \frac{1}{2}(1 + E)$ and $P_{II} = \frac{1}{2}(1 + E)$. In the general case, the probability of guessing a_K correctly is given by:

$$P_K = \frac{1}{2}[1 + E^n] \tag{2.19}$$

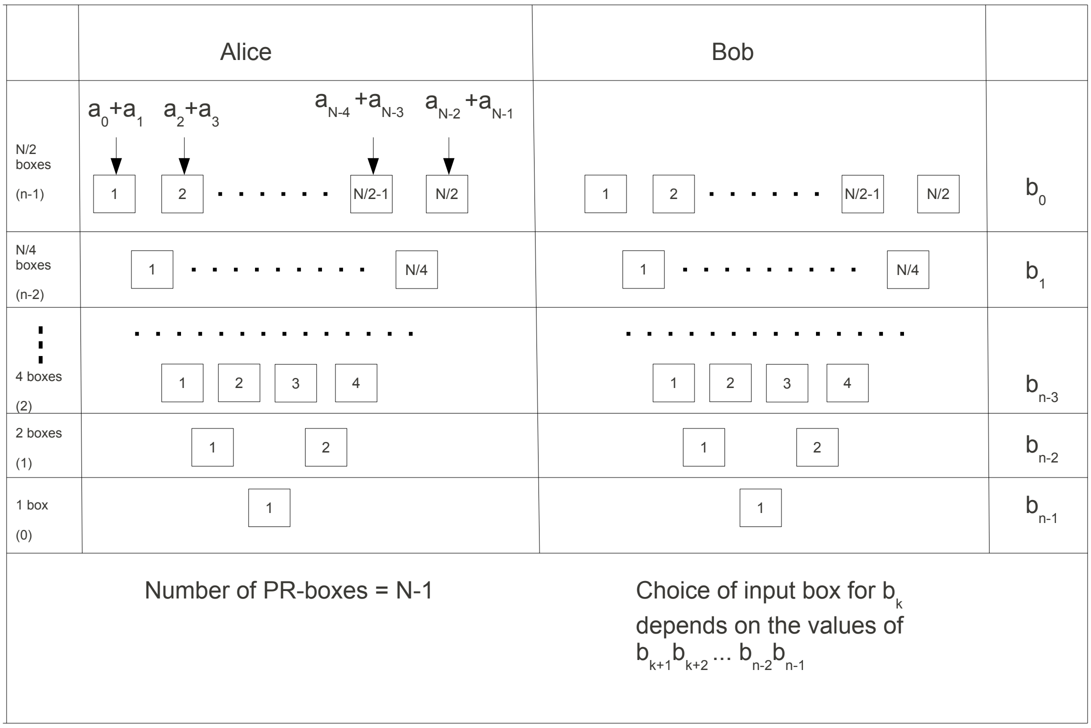

Figure 2.4: The IC game for N bits, with $N-1$ PR-boxes

The accessible information I is:

$$
\begin{aligned}
I &= \sum_{K=0}^{N-1} I(a_K : \beta | b = K) && (2.20)\\
&= \sum_{K=0}^{N-1} (1 - H(a_K|\beta, b = K)) && (2.21)\\
&\geq \sum_{K=0}^{N-1} (1 - h(P_K)) && (2.22)\\
&\geq \sum_{K=0}^{N-1} \frac{E^{2n}}{2\ln 2} && (2.23)\\
&= \frac{(2E^2)^n}{2\ln 2} && (2.24)
\end{aligned}
$$

where we use the inequality $1 - h\left(\frac{1+y}{2}\right) \geq \frac{y^2}{2\ln 2}$. Clearly, for $2E^2 > 1$ or $E > \frac{1}{\sqrt{2}}$ there exist n such than $I > 1$, thus violating information causality. All nonsignalling correlations beyond Tsirelson's bound ($E \leq \frac{1}{\sqrt{2}}$) are therefore ruled out by information causality. One may refer to [8] for the details we have omitted. Also, for a review of information causality and related progress one may refer to [16].

2.4 Macroscopic Locality

Following some work on characterizing quantum correlations numerically [27], Navascués and Wunderlich [18] showed that a physical principle called "macroscopic locality" provides criteria for distinguishing physical correlations from unphysical ones. By "physical" one refers to microscopic correlations that lead to classical behaviour macroscopically. It turns out that the set of such microscopic correlations is not equal to the set of quantum correlations but the former in fact contains the quantum set. We will make this notion precise and point out the relationship between correlations consistent with macroscopic locality and correlations realizable within quantum theory.

2.4.1 Microscopic experiment

A typical microscopic experiment in the Bell scenario consists of two parties, Alice and Bob, receiving a particle each from some common source emitting a pair of particles and subjecting the particle they receive to some interactions X and Y respectively, where $X \in \{1, 2, ..., s\}$ and $Y \in \{s+1, s+2, ..., 2s\}$. Their experimental apparatus consists of detectors that click when a particle arrives and the corresponding outcome is registered. Let's call Alice's outcome a and Bob's outcome b. We denote, following the notation in [18], the detectors corresponding to outcomes a and b by $D(a)$ and $D(b)$ respectively. Also, we denote the respective measurement settings by $X(a)$ and $Y(b)$. $a \in X$ is to be understood as $X(a) = X$, and $b \in Y$ as $Y(b) = Y$. After repeating this experiment a statistically significant number of times, one obtains the probabilities $P(a, b)$ of detector $D(a)$ clicking when interaction $X(a)$ is applied by Alice, and detector $D(b)$ clicking when interaction $Y(b)$ is applied by Bob. The no-signalling condition is assumed to hold, i.e., for any $X \neq X'$, $\sum_{a \in X} P(a, b) = \sum_{a \in X'} P(a, b) \equiv P(b)$, and for any $Y \neq Y'$, $\sum_{b \in Y} P(a, b) = \sum_{b \in Y'} P(a, b) \equiv P(a)$. So choice of measurement settings by one party cannot influence the statistics observed by the other party. See figure (2.5).

The task is to find constraints on $P(a, b)$ from macroscopic locality, i.e., from requiring consistency with classical physics when the experiment is per-

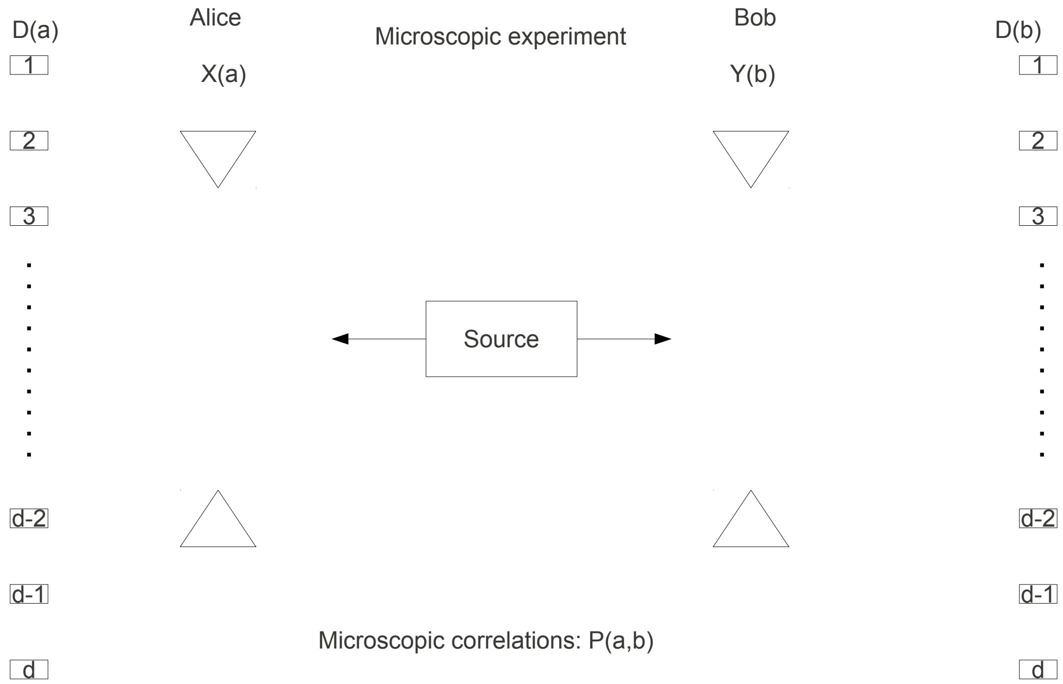

Figure 2.5: Microscopic experiment

formed in a macroscopic setting. It turns out that only a specific set of microscopic correlations $P(a, b)$ can yield a classical model at the macroscopic level.

2.4.2 Macroscopic experiment

In a typical macroscopic experiment, a common source emits N independent pairs of particles, where $N \gg 1$. Alice and Bob will each receive a beam of particles. The interactions (measurement settings) they apply will be applied to all the particles of the respective beam – they do not have access to individual particles. Depending on the interaction Alice (Bob) applies, the initial beam from the source splits up into beams of different intensities received at Alice's (Bob's) corresponding detectors. The measurement outcomes in this macroscopic experiment are not the clicks of individual detectors since all the detectors click in each run of the experiment. Instead, the measurement outcome is the distribution of intensities observed in the detectors in a given run of the experiment. Further, to relate this to classical physics we regard the intensities as continuous fields rather than fluxes of discrete particles. Given this, one needs to assume some resolution for Alice's and Bob's detectors. See figure (2.6).

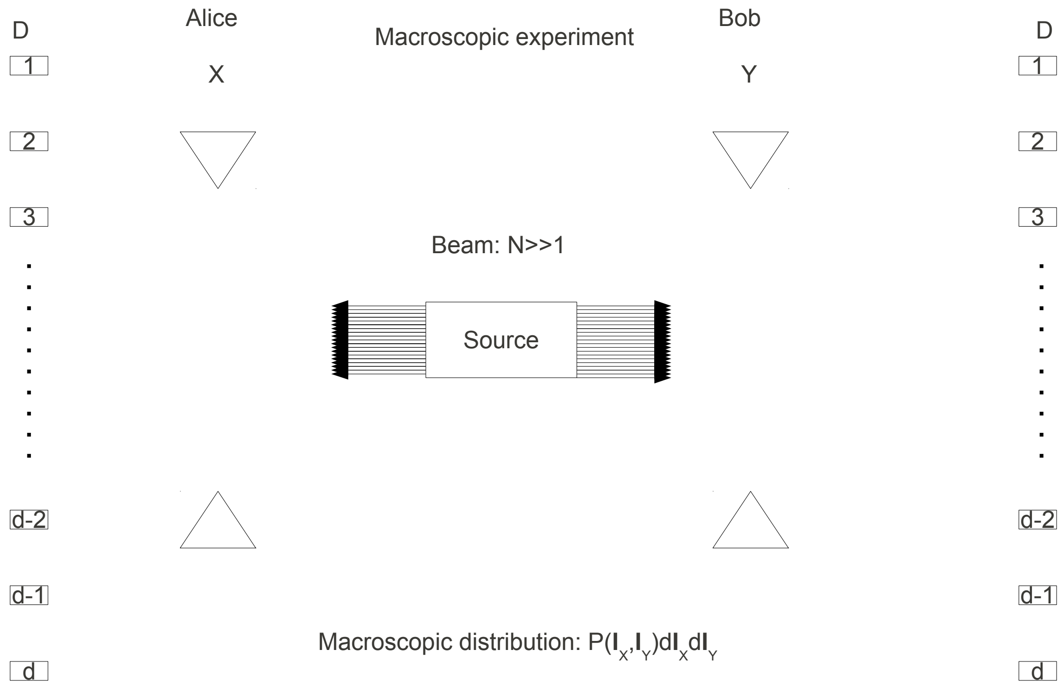

Figure 2.6: Macroscopic experiment

Poor detector resolution will mean that Alice and Bob will always observe the same outcomes in repeated experiments – the intensity reading at detector $D(c)$ being $NP(c)$. The minimum resolution that Alice and Bob need to observe intensity fluctuations about the mean value $NP(c)$ (over different repetitions of the experiment) is of the order $\sqrt{N}$. We will assume this resolution.

2.4.3 Macroscopic locality

In the macroscopic experiment we denote Alice's (Bob's) outcomes by $\mathbf{I}_X$ ($\mathbf{I}_Y$), the intensities observed at d detectors. So $\mathbf{I}_X = (I_X^1, I_X^2, ..., I_X^d)$ and $\mathbf{I}_Y = (I_Y^1, I_Y^2, ..., I_Y^d)$. After many repetitions of the experiment, Alice and Bob can estimate the marginal probability densities $P(\mathbf{I}_X, \mathbf{I}_Y)d\mathbf{I}_X d\mathbf{I}_Y$ for any pair of respective measurements X and Y. Since classical physics is a local theory (in the Bell-local sense), it admits a local hidden variable model. This requires the existence of a *global* probability density $P(\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_{2s})$, the marginals of which yield $P(\mathbf{I}_X, \mathbf{I}_Y)$:

$$P(\mathbf{I}_X, \mathbf{I}_Y) = \int \left(\prod_{Z \neq X,Y} d\mathbf{I}_Z \right) P(\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_{2s}) \tag{2.25}$$

This is the constraint of *macroscopic locality.*

Characterizing macroscopic locality: In the bipartite case, the set of microscopic correlations that yield macroscopic distributions satisfying macroscopic locality is equal to the set Q^1 defined in [27] via semidefinite programming as a first approximation to the set of quantum correlations. It turns out that the set of bipartite quantum correlations Q is contained within the set of correlations consistent with macroscopic locality, i.e., $Q \subset Q^1$. In fact, quantum correlations are macroscopically local even in the multipartite case. Clearly, there exist correlations not realizable within quantum theory that are nevertheless consistent with macroscopic locality. One can think of macroscopic locality as a test that lets one discard "unphysical" modifications of quantum theory – those that predict correlations violating macroscopic locality. We briefly sketch the proof from [18]:

A. The set of macroscopically local correlations is equal to Q^1: To begin with, we need to identify the relationship between the microscopic distributions $P(a, b)$ and the corresponding macroscopic probability densities $P(\mathbf{I}_X, \mathbf{I}_Y)$. Recall that a (b) denotes click of detector $D(a)$ ($D(b)$) when Alice (Bob) applies interaction $X(a)$ ($Y(b)$). In the macroscopic experiment there are N pairs of particles, $N \gg 1$. Consider a binary valued variable d_i^c:

$$d_i^c = \left\{ \begin{array}{ll} 1, & \text{if a particle in the } i\text{th pair arrives at detector } c \\ 0, & \text{otherwise} \end{array} \right\}$$

Now, the intensity observed in a detector c, I^c, is proportional to $\sum_{i=1}^{N} d_i^c$. The probability that detector c clicks in a given run of the macroscopic experiment is given by

$$P(c) = \frac{\sum_{i=1}^{N} d_i^c}{N} = \langle d_i^c \rangle$$

Since we are interested in intensity fluctuations about the mean, we define

$$\bar{d}_i^c = d_i^c - P(c)$$

Clearly,

$$\langle \bar{d}_i^c \rangle = \langle d_i^c \rangle - P(c) = P(c) - P(c) = 0$$

So we have $\langle \bar{d}_i^a \rangle = \langle \bar{d}_i^b \rangle = 0$, $\forall\ X(a) = X, Y(b) = Y$. The fluctuation in intensity observed in detector c is thus given by

$$\bar{I}^c \equiv \sum_{i=1}^{N} \frac{\bar{d}_i^c}{\sqrt{N}} \tag{2.26}$$

The factor of $1/\sqrt{N}$ takes into account the resolution we have assumed for Alice's and Bob's detectors. From the central limit theorem it follows that in the limit $N \to \infty$, the probability distribution governing the fluctuations $\bar{I}^c$ converges to a multivariate gaussian distribution of mean $\vec{0}$ and covariance matrix γ^{XY} given by:

$$\begin{aligned}
\gamma_{cc'}^{XY} &= \langle \bar{I}^c \bar{I}^{c'} \rangle && (2.27)\\
&= \frac{1}{N} \sum_{i,j=1}^{N} \langle \bar{d}_i^c \bar{d}_j^{c'} \rangle && (2.28)\\
&= \frac{1}{N} \Big(\sum_{i=1}^{N} \langle \bar{d}_i^c \bar{d}_i^{c'} \rangle + \sum_{i \neq j} \langle \bar{d}_i^c \bar{d}_j^{c'} \rangle \Big) && (2.29)\\
&= \langle \bar{d}_1^c \bar{d}_1^{c'} \rangle && (2.30)\\
&= P(c, c') - P(c)P(c') && (2.31)
\end{aligned}$$

where we have used the fact that the N pairs are identical and uncorrelated. The matrix γ^{XY} is a $2d \times 2d$ matrix, where d is the number of detectors on either side. Obviously, $c, c' \in \{(X, D), (Y, D) | D \in \{1, 2, ..., d\}\}$. X (Y) labels the interaction applied by Alice (Bob) and D the detector corresponding to the outcome. This covariance matrix looks like the following:

$$\gamma^{XY} = \begin{pmatrix}
\langle \bar{I}^{(X,1)} \bar{I}^{(X,1)} \rangle & \cdots & \langle \bar{I}^{(X,1)} \bar{I}^{(X,d)} \rangle & \langle \bar{I}^{(X,1)} \bar{I}^{(Y,1)} \rangle & \cdots & \langle \bar{I}^{(X,1)} \bar{I}^{(Y,d)} \rangle \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\langle \bar{I}^{(X,d)} \bar{I}^{(X,1)} \rangle & \cdots & \langle \bar{I}^{(X,d)} \bar{I}^{(X,d)} \rangle & \langle \bar{I}^{(X,d)} \bar{I}^{(Y,1)} \rangle & \cdots & \langle \bar{I}^{(X,d)} \bar{I}^{(Y,d)} \rangle \\
\langle \bar{I}^{(Y,1)} \bar{I}^{(X,1)} \rangle & \cdots & \langle \bar{I}^{(Y,1)} \bar{I}^{(X,d)} \rangle & \langle \bar{I}^{(Y,1)} \bar{I}^{(Y,1)} \rangle & \cdots & \langle \bar{I}^{(Y,1)} \bar{I}^{(Y,d)} \rangle \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\langle \bar{I}^{(Y,d)} \bar{I}^{(X,1)} \rangle & \cdots & \langle \bar{I}^{(Y,d)} \bar{I}^{(X,d)} \rangle & \langle \bar{I}^{(Y,d)} \bar{I}^{(Y,1)} \rangle & \cdots & \langle \bar{I}^{(Y,d)} \bar{I}^{(Y,d)} \rangle
\end{pmatrix}$$

We have shown how the set of microscopic correlations produces marginal gaussian probability distributions in the macroscopic limit $N \gg 1$. The condition of macroscopic locality requires that this set of gaussian marginals should arise from a global joint probability distribution. For such a global probability distribution to exist, the corresponding *global* covariance matrix Γ should be positive semidefinite, with $\Gamma_{cc'} = \langle \bar{I}^c \bar{I}^{c'} \rangle$, for all $c, c' \in \{(Z, D)|Z \in \{1, 2, ..., 2s\}, D \in \{1, 2, ..., d\}\}$. (Recall that Alice has s available settings $X \in \{1, 2, ..., s\}$, and Bob has s settings $Y \in \{s+1, ..., 2s\}$.)

Thus we have $\Gamma \geq 0$ of the form

$$\Gamma = \begin{pmatrix} Q & P \\ P^T & R \end{pmatrix} \tag{2.32}$$

where Q, P, and R are the following matrices:

$$Q = \begin{pmatrix}
\langle \bar{I}^{(1,1)} \bar{I}^{(1,1)} \rangle & \cdots & \langle \bar{I}^{(1,1)} \bar{I}^{(1,d)} \rangle & \cdots & \langle \bar{I}^{(1,1)} \bar{I}^{(s,1)} \rangle & \cdots & \langle \bar{I}^{(1,1)} \bar{I}^{(s,d)} \rangle \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\langle \bar{I}^{(1,d)} \bar{I}^{(1,1)} \rangle & \cdots & \langle \bar{I}^{(1,d)} \bar{I}^{(1,d)} \rangle & \cdots & \langle \bar{I}^{(1,d)} \bar{I}^{(s,1)} \rangle & \cdots & \langle \bar{I}^{(1,d)} \bar{I}^{(s,d)} \rangle \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\langle \bar{I}^{(s,1)} \bar{I}^{(1,1)} \rangle & \cdots & \langle \bar{I}^{(s,1)} \bar{I}^{(1,d)} \rangle & \cdots & \langle \bar{I}^{(s,1)} \bar{I}^{(s,1)} \rangle & \cdots & \langle \bar{I}^{(s,1)} \bar{I}^{(s,d)} \rangle \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\langle \bar{I}^{(s,d)} \bar{I}^{(1,1)} \rangle & \cdots & \langle \bar{I}^{(s,d)} \bar{I}^{(1,d)} \rangle & \cdots & \langle \bar{I}^{(s,d)} \bar{I}^{(s,1)} \rangle & \cdots & \langle \bar{I}^{(s,d)} \bar{I}^{(s,d)} \rangle
\end{pmatrix}$$

That is, Q is an $sd \times sd$ matrix with rows and columns labelled by Alice's measurement settings and outcomes $\{(X, D)|X \in \{1, ..., s\}, D \in \{1, ..., d\}\}$.

$$R = \begin{pmatrix}
\langle \bar{I}^{(s+1,1)} \bar{I}^{(s+1,1)} \rangle & \cdots & \langle \bar{I}^{(s+1,1)} \bar{I}^{(s+1,d)} \rangle & \cdots & \langle \bar{I}^{(s+1,1)} \bar{I}^{(2s,1)} \rangle & \cdots & \langle \bar{I}^{(s+1,1)} \bar{I}^{(2s,d)} \rangle \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\langle \bar{I}^{(s+1,d)} \bar{I}^{(s+1,1)} \rangle & \cdots & \langle \bar{I}^{(s+1,d)} \bar{I}^{(s+1,d)} \rangle & \cdots & \langle \bar{I}^{(s+1,d)} \bar{I}^{(2s,1)} \rangle & \cdots & \langle \bar{I}^{(s+1,d)} \bar{I}^{(2s,d)} \rangle \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\langle \bar{I}^{(2s,1)} \bar{I}^{(s+1,1)} \rangle & \cdots & \langle \bar{I}^{(2s,1)} \bar{I}^{(s+1,d)} \rangle & \cdots & \langle \bar{I}^{(2s,1)} \bar{I}^{(2s,1)} \rangle & \cdots & \langle \bar{I}^{(2s,1)} \bar{I}^{(2s,d)} \rangle \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\langle \bar{I}^{(2s,d)} \bar{I}^{(s+1,1)} \rangle & \cdots & \langle \bar{I}^{(2s,d)} \bar{I}^{(s+1,d)} \rangle & \cdots & \langle \bar{I}^{(2s,d)} \bar{I}^{(2s,1)} \rangle & \cdots & \langle \bar{I}^{(2s,d)} \bar{I}^{(2s,d)} \rangle
\end{pmatrix}$$

That is, R is an $sd \times sd$ matrix with rows and columns labelled by Bob's measurement settings and outcomes $\{(Y, D)|Y \in \{s+1, ..., 2s\}, D \in \{1, ..., d\}\}$.

$$P = \begin{pmatrix} \langle \bar{I}^{(1,1)} \bar{I}^{(s+1,1)} \rangle & \dots & \langle \bar{I}^{(1,1)} \bar{I}^{(s+1,d)} \rangle & \dots & \langle \bar{I}^{(1,1)} \bar{I}^{(2s,1)} \rangle & \dots & \langle \bar{I}^{(1,1)} \bar{I}^{(2s,d)} \rangle \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle \bar{I}^{(1,d)} \bar{I}^{(s+1,1)} \rangle & \dots & \langle \bar{I}^{(1,d)} \bar{I}^{(s+1,d)} \rangle & \dots & \langle \bar{I}^{(1,d)} \bar{I}^{(2s,1)} \rangle & \dots & \langle \bar{I}^{(1,d)} \bar{I}^{(2s,d)} \rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle \bar{I}^{(s,1)} \bar{I}^{(s+1,1)} \rangle & \dots & \langle \bar{I}^{(s,1)} \bar{I}^{(s+1,d)} \rangle & \dots & \langle \bar{I}^{(s,1)} \bar{I}^{(2s,1)} \rangle & \dots & \langle \bar{I}^{(s,1)} \bar{I}^{(2s,d)} \rangle \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle \bar{I}^{(s,d)} \bar{I}^{(s+1,1)} \rangle & \dots & \langle \bar{I}^{(s,d)} \bar{I}^{(s+1,d)} \rangle & \dots & \langle \bar{I}^{(s,d)} \bar{I}^{(2s,1)} \rangle & \dots & \langle \bar{I}^{(s,d)} \bar{I}^{(2s,d)} \rangle \end{pmatrix}$$

That is, P is an $sd \times sd$ matrix with rows labelled by Alice's measurement settings and outcomes $\{(X, D)|X \in \{1, ..., s\}, D \in \{1, ..., d\}\}$ and columns labelled by Bob's measurement settings and outcomes $\{(Y, D)|Y \in \{s + 1, ..., 2s\}, D \in \{1, ..., d\}\}$. The matrix elements of P are given by:

$$P_{ab} = P(a, b) - P(a)P(b) \tag{2.33}$$

Q and R are determined only partially:

$$Q_{a,a'} = \delta_{aa'} P(a) - P(a)P(a'), \text{ if } X(a) = X(a') \tag{2.34}$$
$$R_{b,b'} = \delta_{bb'} P(b) - P(b)P(b'), \text{ if } Y(b) = Y(b') \tag{2.35}$$

The remaining matrix elements of Q and R, $\{Q_{a,a'}, R_{b,b'}|X(a) \neq X(a'), Y(b) \neq Y(b')\}$, correspond to correlations between different settings and are therefore not measurable directly. However, for the gaussian marginals to arise from a global distribution they should take values such that $\Gamma \geq 0$. Conversely, if such matrix elements exist then there exists a global probability distribution with mean $\vec{0}$ and covariance matrix Γ. This gives us a set of necessary and sufficient conditions for a set of microscopic correlations $P(a, b)$ to be macroscopically local. Having thus characterized the set of macroscopically local correlations, we turn to the set of correlations Q^1 [27].

Navascués et al. [27] defined Q^1 as follows: $P(a, b) \in Q^1$ if and only if there exists a positive semidefinite matrix γ (i.e., $\gamma \geq 0$) of the form

$$\gamma = \begin{pmatrix} 1 & \vec{P}_A^T & \vec{P}_B^T \\ \vec{P}_A & \tilde{Q} & \tilde{P}^T \\ \vec{P}_B & \tilde{P} & \tilde{R} \end{pmatrix} \tag{2.36}$$

where $\vec{P}_A = (P_1(1), ..., P_1(d), ..., P_s(1), ..., P_s(d))$ is the $sd \times 1$ column vector of probabilities $P_{X(a)}(a)$ for Alice (the subscript denotes the measurement setting). Similarly, $\vec{P}_B = (P_{s+1}(1), ..., P_{s+1}(d), ..., P_{2s}(1), ..., P_{2s}(d))$ is the $sd \times 1$ column vector of probabilities $P_{Y(b)}(b)$ for Bob. $\tilde{Q}$ is a $sd \times sd$ matrix of the form

$$\tilde{Q} = \begin{pmatrix} \tilde{Q}_{(1,1)(1,1)} & \cdots & \tilde{Q}_{(1,1)(1,d)} & \cdots & \tilde{Q}_{(1,1)(s,1)} & \cdots & \tilde{Q}_{(1,1)(s,d)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{Q}_{(1,d)(1,1)} & \cdots & \tilde{Q}_{(1,d)(1,d)} & \cdots & \tilde{Q}_{(1,d)(s,1)} & \cdots & \tilde{Q}_{(1,d)(s,d)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{Q}_{(s,1)(1,1)} & \cdots & \tilde{Q}_{(s,1)(1,d)} & \cdots & \tilde{Q}_{(s,1)(s,1)} & \cdots & \tilde{Q}_{(s,1)(s,d)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{Q}_{(s,d)(1,1)} & \cdots & \tilde{Q}_{(s,d)(1,d)} & \cdots & \tilde{Q}_{(s,d)(s,1)} & \cdots & \tilde{Q}_{(s,d)(s,d)} \end{pmatrix}$$

labelled by Alice's measurement outcomes. The matrix elements are partially determined by:

$$\tilde{Q}_{aa'} = \delta_{aa'} P(a) \text{ if } X(a) = X(a') \tag{2.37}$$

Similarly,

$$\tilde{R} = \begin{pmatrix} \tilde{R}_{(s+1,1)(s+1,1)} & \cdots & \tilde{R}_{(s+1,1)(s+1,d)} & \cdots & \tilde{R}_{(s+1,1)(2s,1)} & \cdots & \tilde{R}_{(s+1,1)(2s,d)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{R}_{(s+1,d)(s+1,1)} & \cdots & \tilde{R}_{(s+1,d)(s+1,d)} & \cdots & \tilde{R}_{(s+1,d)(2s,1)} & \cdots & \tilde{R}_{(s+1,d)(2s,d)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{R}_{(2s,1)(s+1,1)} & \cdots & \tilde{R}_{(2s,1)(s+1,d)} & \cdots & \tilde{R}_{(2s,1)(2s,1)} & \cdots & \tilde{R}_{(2s,1)(2s,d)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{R}_{(2s,d)(s+1,1)} & \cdots & \tilde{R}_{(2s,d)(s+1,d)} & \cdots & \tilde{R}_{(2s,d)(2s,1)} & \cdots & \tilde{R}_{(2s,d)(2s,d)} \end{pmatrix}$$

where

$$\tilde{R}_{bb'} = \delta_{bb'} P(b) \text{ if } Y(b) = Y(b') \tag{2.38}$$

$\tilde{P}$ is a matrix with rows labelled by Bob's measurement outcomes and columns labelled by Alice's measurement outcomes:

$$\tilde{P} = \begin{pmatrix} \tilde{P}_{(s+1,1)(1,1)} & \cdots & \tilde{P}_{(s+1,1)(1,d)} & \cdots & \tilde{P}_{(s+1,1)(s,1)} & \cdots & \tilde{P}_{(s+1,1)(s,d)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{P}_{(s+1,d)(1,1)} & \cdots & \tilde{P}_{(s+1,d)(1,d)} & \cdots & \tilde{P}_{(s+1,d)(s,1)} & \cdots & \tilde{P}_{(s+1,d)(s,d)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{P}_{(2s,1)(1,1)} & \cdots & \tilde{P}_{(2s,1)(1,d)} & \cdots & \tilde{P}_{(2s,1)(s,1)} & \cdots & \tilde{P}_{(2s,1)(s,d)} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{P}_{(2s,d)(1,1)} & \cdots & \tilde{P}_{(2s,d)(1,d)} & \cdots & \tilde{P}_{(2s,d)(s,1)} & \cdots & \tilde{P}_{(2s,d)(s,d)} \end{pmatrix}$$

where

$$\tilde{P}_{ba} = P(a,b) \tag{2.39}$$

Given this characterization of Q^1 we can now show that it is equivalent to the set of macroscopically local correlations using Schur's theorem.

Schur's theorem: Let H be a matrix of the form

$$H = \begin{pmatrix} E & F \\ F^T & G \end{pmatrix} \tag{2.40}$$

such that $E > 0$. Then $H \geq 0$ iff $G - F^T E^{-1} F \geq 0$.

Applying this theorem to γ (cf. 2.36), we identify $H \equiv \gamma$, $E \equiv 1$, $F \equiv (\vec{P}_A^T, \vec{P}_B^T)$, and

$$G \equiv \begin{pmatrix} \tilde{Q} & \tilde{P}^T \\ \tilde{P} & \tilde{R} \end{pmatrix} \tag{2.41}$$

$E > 0$. Therefore, the condition for $\gamma \geq 0$ becomes

$$\begin{pmatrix} \tilde{Q} & \tilde{P}^T \\ \tilde{P} & \tilde{R} \end{pmatrix} - \begin{pmatrix} \vec{P}_A \\ \vec{P}_B \end{pmatrix} . \begin{pmatrix} \vec{P}_A^T & \vec{P}_B^T \end{pmatrix} \geq 0 \tag{2.42}$$

This matrix is in fact equal to Γ (cf. 2.32), i.e.,

$$\Gamma = \begin{pmatrix} \tilde{Q} - \vec{P}_A.\vec{P}_A^T & \tilde{P}^T - \vec{P}_A.\vec{P}_B^T \\ \tilde{P} - \vec{P}_B.\vec{P}_A^T & \tilde{R} - \vec{P}_B.\vec{P}_B^T \end{pmatrix} \tag{2.43}$$

where $\tilde{Q} - \vec{P}_A.\vec{P}_A^T = Q$, $\tilde{P}^T - \vec{P}_A.\vec{P}_B^T = P$, $\tilde{P} - \vec{P}_B.\vec{P}_A^T = P^T$, and $\tilde{R} - \vec{P}_B.\vec{P}_B^T = R$. Of course, this requires one to change the undetermined coefficients according to:

$$Q_{aa'} = \tilde{Q}_{aa'} - P(a)P(a'), \text{ if } X(a) \neq X(a') \tag{2.44}$$
$$R_{bb'} = \tilde{R}_{bb'} - P(b)P(b'), \text{ if } Y(b) \neq Y(b') \tag{2.45}$$

Thus, $\gamma \geq 0$ if and only if $\Gamma \geq 0$. The set Q^1 and the set of macroscopically local correlations are equivalent.

B. Quantum correlations satisfy macroscopic locality. Consider a self-adjoint operator $\tilde{c}$ associated with each gaussian variable c (denoting the measurement setting and the detector corresponding to the outcome). We define $\Gamma_{cc'} \equiv \mathrm{Tr}(\rho[\tilde{c} - \langle c\rangle \mathbb{I}, \tilde{c}' - \langle c'\rangle \mathbb{I}]_+)/2$, where the anti-commutator

$[\tilde{c}-\langle c\rangle\mathbb{I},\tilde{c}'-\langle c'\rangle\mathbb{I}]_+ = (\tilde{c}-\langle c\rangle\mathbb{I})(\tilde{c}'-\langle c'\rangle\mathbb{I})+(\tilde{c}'-\langle c'\rangle\mathbb{I})(\tilde{c}-\langle c\rangle\mathbb{I})$. We therefore have a real matrix with $\Gamma_{cc'} = \langle (c-\langle c\rangle)(c'-\langle c'\rangle)\rangle$ for any pair of compatible variables c, c' (i.e., $[\tilde{c},\tilde{c}'] = 0$). Positive semidefiniteness of this real matrix implies that it corresponds to a gaussian LHV model macroscopically, i.e., the set of quantum correlations satisfy macroscopic locality. For an arbitrary real vector $\vec{v}$,

$$
\begin{aligned}
\vec{v}^T\Gamma\vec{v} &= \sum_c (\vec{v}^T\Gamma\vec{v})_c && (2.46)\\
&= \sum_c (\sum_{c'} v_c\Gamma_{cc'}v_{c'}) && (2.47)\\
&= \mathrm{Tr}(\rho(\sum_c v_c(\tilde{c}-\langle c\rangle\mathbb{I}))(\sum_{c'} v_{c'}(\tilde{c}'-\langle c'\rangle\mathbb{I}))) && (2.48)\\
&= \mathrm{Tr}(\rho M M^\dagger) && (2.49)\\
&\geq 0 && (2.50)
\end{aligned}
$$

where $M = M^\dagger = \sum_c v_c(\tilde{c}-\langle c\rangle\mathbb{I})$. Since $\vec{v}$ is arbitrary, $\Gamma \geq 0$. Therefore, a multipartite quantum mechanical state ρ gives rise to macroscopic intensity fluctuations which admit a local hidden variable model, i.e., quantum correlations are macroscopically local, $Q \subset Q^1$.

C. The set Q^1 is closed under wiring. Consider a set of n no-signalling boxes producing correlations $P_i(a,b) \in Q^1, i \in \{1,2,...,n\}$, where each box is shared between Alice and Bob. Alice (Bob) can make local measurements on her (his) part of the shared boxes in any order she (he) chooses and can "wire" the boxes together, e.g., feed the output from one box as an input to another, and at the end of this "wiring" protocol this set of no-signalling boxes (treated as one single box) produces an output on both Alice's side and Bob's side, given some inputs on either side. We label the resulting correlations by $P(a,b)$. The question is: Do the correlations that arise from this wiring obey macroscopic locality, i.e., is $P(a,b) \in Q^1$? It turns out that they do – the set Q^1 is closed under any such wiring. Wiring cannot generate correlations violating macroscopic locality from boxes producing macroscopically local correlations. (See [18] for the proof.)

D. All two-point correlators accessible in Q^1 admit a quantum representation. Consider the two-point correlator E_{XY}:

$$E_{XY} \equiv \langle O_X O_Y\rangle = \sum_{a,b\in X,Y} P(a,b)O_X(a)O_Y(b) \tag{2.51}$$

In the CHSH scenario, we have $s = d = 2$, so our measurement settings are labelled by $1, 2$ (Alice) and $3, 4$ (Bob). The outcomes are assigned real values ± 1 according to $O_Z : c \to \{1, -1\}$, $Z \in \{1, 2, 3, 4\}$. The $\langle CHSH \rangle$ parameter is defined by

$$\langle CHSH \rangle \equiv E_{13} + E_{14} + E_{23} - E_{24}.$$

Local correlations obey $|\langle CHSH \rangle| \leq 2$, and quantum correlations satisfy $|\langle CHSH \rangle| \leq 2\sqrt{2}$ (Tsirelson's bound). It turns out that $|\langle CHSH \rangle| \leq 2\sqrt{2}$ for the set of macroscopically local correlations Q^1. Indeed, a macroscopically local theory gives rise to correlators $\{E_{13}, E_{14}, E_{23}, E_{24}\}$ that admit a quantum mechanical model ([27]). This, in fact, holds for arbitrary s with $d = 2$. However, the set of two-point correlators E_{XY} does not contain all the information about the distribution $P(a, b)$, so the quantum set and the macroscopically local set are not identical. Quantum correlations lie strictly within the set of macroscopically local correlations, i.e., $Q \subset Q^1$, even for the $s = d = 2$ case.

2.5 Inequivalence of Information Causality and Macroscopic Locality

We have seen two principles that could characterize the set of quantum correlations – macroscopic locality (ML) and information causality (IC). A natural question that arises is whether these two principles are compatible with each other. In this section we mention a result [28] that shows the existence of macroscopically local correlations that violate information causality. IC and ML are therefore inequivalent principles.

The IC scheme is first generalized to dits (or d-level inputs with modulo d arithmetic) instead of the original $d = 2$ case with bits. Alice receives a string $\vec{a}$ of N dits, Bob receives the index b of the dit to be retrieved. Alice and Bob use nonsignalling boxes modelled as noisy PR-boxes with E denoting the noise level. Now, the critical noise level beyond which IC is violated is denoted by E_{IC}, and the noise level beyond which ML is violated is denoted by E_{ML}. For $d = 2$, $E_{IC} = E_{ML} = E_Q = 0.707$, where $E_Q = \frac{1}{\sqrt{2}}$ is the quantum (Tsirelson's) bound. For $d = 5$, we have $E_{IC} \lesssim 0.700$ and $E_{ML} = 0.707$. Clearly, for noise level between 0.700 and 0.707, IC is violated while ML is satisfied. Thus, there exist correlations consistent with macroscopic locality that violate information causality. The two principles are therefore inequivalent.

2.6 Quantum nonlocality and the Uncertainty principle

In [19], Oppenheim and Wehner demonstrated a connection between nonlocality and the uncertainty principle. The main conclusion of their work is that the strength of the uncertainty principle, together with the ability to remotely prepare states by making local measurements (known as **steering**), limits the strength of nonlocal correlations. In the case of quantum theory, this limit imposed by the uncertainty principle yields the well-known Tsirelson's bound. Their result, however, is more generally valid in theories with convex state spaces, where correlations stronger than quantum are in general allowed. The degree of nonlocality manifest in any theory, according to this work, is determined by a tradeoff between uncertainty relations and steerability in the theory. Strong uncertainty relations and limited steering restrict the strength of nonlocal correlations. The uncertainty relations in quantum theory, together with its steering properties, restrict quantum correlations to Tsirelson's bound. We discuss this work in this section.

We will review the main result of [19], introducing fine-grained uncertainty relations (fgURs), steering, and the interplay between the two in the bipartite Bell scenario where nonlocality is shown to be constrained by fgURs and steering constraints.

2.6.1 Fine-grained Uncertainty Relations (fgURs)

Consider a single measurement, denoted by t, that is made on a system in some state σ. Let $x^{(t)}$ denote the measurement outcome which can take one of a set of possible values, according to the distribution $p(x^{(t)}|t)_\sigma$. Typically, entropic uncertainty relations [26] involve a coarse-graining over these probabilities, for example, in calculating the Shannon entropy associated with (t, σ):

$$H(t)_\sigma = -\sum_{x^{(t)}} p(x^{(t)}|t)_\sigma \log p(x^{(t)}|t)_\sigma$$

An entropic uncertainty relation – capturing the average entropy of a set of measurements $\mathcal{T} = \{1, 2, ..., n\}$ – would then be of the type:

$$\sum_{t\in\mathcal{T}} p(t)H(t)_\sigma \geq c_{\mathcal{T},\mathcal{D}}$$

where the measurements are chosen according to the probability distribution $\mathcal{D} = \{p(t)\}_{t\in\mathcal{T}}$ and $c_{\mathcal{T},\mathcal{D}} > 0$.

Fine-grained uncertainty relations, on the other hand, do not involve a coarse-graining of this kind. Consider the set of measurements $\mathcal{T} = \{1, 2, ..., n\}$, where the outcomes for each $t \in \mathcal{T}$ take values $x^{(t)} \in \mathcal{B}$. Consider also a string of outcomes $\vec{x} = (x^{(1)}, x^{(2)}, ..., x^{(n)})$ for all the measurements $t \in \mathcal{T}$. Clearly, $\vec{x} \in \mathcal{B}^{\times n}$. The measurements are performed according to $\mathcal{D} = \{p(t)\}_{t \in \mathcal{T}}$. For a fixed choice of $(\mathcal{T}, \mathcal{D})$, the set of inequalities

$$\mathcal{U} = \{\sum_{t=1}^{n} p(t)p(x^{(t)}|t)_\sigma \leq \zeta_{\vec{x}} \quad | \quad \forall \vec{x} \in \mathcal{B}^{\times n}\} \tag{2.52}$$

forms a fine-grained uncertainty relation (fgUR). Here, for a given $\vec{x} \in \mathcal{B}^{\times n}$

$$\zeta_{\vec{x}} \equiv \max_\sigma \sum_{t=1}^{n} p(t)p(x^{(t)}|t)_\sigma = \sum_{t=1}^{n} p(t)p(x^{(t)}|t)_{\rho_{\vec{x}}} \tag{2.53}$$

where the maximization is taken over all allowed states σ and $\rho_{\vec{x}}$ is the "maximally certain state" (or the minimal uncertainty state) that saturates the corresponding fgUR inequality. Also, $0 \leq \zeta_{\vec{x}} \leq 1 \quad \forall \vec{x} \in \mathcal{B}^{\times n}$, where:

1. $\zeta_{\vec{x}} = 0 \Rightarrow$ outcomes in string $\vec{x}$ do not occur in any measurement, i.e., $p(x^{(t)}|t)_\sigma = 0 \quad \forall x^{(t)} \in \vec{x}$, for all allowed states σ.

2. $0 < \zeta_{\vec{x}} < 1 \Rightarrow p(x^{(t)}|t)_\sigma < 1$ for at least one $x^{(t)} \in \vec{x}$ for any allowed state σ.

3. $\zeta_{\vec{x}} = 1 \Rightarrow p(x^{(t)}|t)_\sigma = 1, \quad \forall x^{(t)} \in \vec{x}$, for the maximally certain state $\rho_{\vec{x}}$.

Clearly (cf. 2.52), the set of values $\{\zeta_{\vec{x}}\}_{\vec{x} \in \mathcal{B}^{\times n}}$ constrains the statistics of measurement outcomes in the theory, thus embodying the uncertainty principle.

2.6.2 Nonlocal correlations

We consider the CHSH Bell scenario, $(2, 2, 2)$, i.e., two parties (Alice and Bob), two measurements available to each, and two possible outcomes for each measurement (cf. 1.3). We have described the CHSH game earlier (1.3) and here we merely recast it in a form amenable to employing fgURs. In the CHSH game, Alice and Bob share a no-signalling resource prepared in a certain state. Alice receives a question $s \in \{0, 1\}$ and Bob receives a question $t \in \{0, 1\}$. When Alice receives s she performs the corresponding measurement on her part of the shared state, notes the outcome, and accordingly answers $a \in \{0,1\}$. The measurement-outcome pair (s,a) of Alice specifies the 'winning string' in the CHSH game: $\vec{x}_{s,a} = (x_{s,a}^{(0)}, x_{s,a}^{(1)})$. Here $x_{s,a}^{(t)}$ is the answer Bob must produce if he receives a question $t \in \{0,1\}$ for Alice and Bob to win the game. According to the winning conditions we know (cf. 1.12), Alice and Bob must produce identical answers except when they are asked questions $s = 1, t = 1$. Therefore, the 'winning strings' are:

$$\vec{x}_{0,0} = (0,0) \quad (2.54)$$
$$\vec{x}_{0,1} = (1,1) \quad (2.55)$$
$$\vec{x}_{1,0} = (0,1) \quad (2.56)$$
$$\vec{x}_{1,1} = (1,0) \quad (2.57)$$

Here, $x_{s,a}^{(t)} = b$ such that $a \oplus b = st$. Alice and Bob are allowed to choose a pre-established strategy, $(\mathcal{S}, \mathcal{T}, \sigma_{AB})$, before the game starts. $\mathcal{S}$ is the set of measurements Alice chooses from, labelled by 0 and 1 in the CHSH scenario. $\mathcal{T}$ is the set of measurements Bob chooses from, labelled by 0 and 1 in the CHSH scenario. And σ_{AB} is a shared state between Alice and Bob. During the game, Alice and Bob receive questions $s \in \mathcal{S}$ and $t \in \mathcal{T}$ respectively, according to a probability distribution $p(s,t)$. The conditional probability of obtaining a winning combination of outcomes, given (s, t, σ_{AB}), is $p(a, b = x_{s,a}^{(t)}|s,t)_{\sigma_{AB}}$. Thus the average winning probability for a particular strategy $(\mathcal{S}, \mathcal{T}, \sigma_{AB})$ is:

$$P^{\text{game}}(\mathcal{S}, \mathcal{T}, \sigma_{AB}) = \sum_{s,t} p(s,t) \sum_{a} p(a, b = x_{s,a}^{(t)}|s,t)_{\sigma_{AB}} \quad (2.58)$$

The strength of correlations is quantified by maximizing this winning probability over all pre-established strategies $(\mathcal{S}, \mathcal{T}, \sigma_{AB})$:

$$P_{\max}^{\text{game}} = \max_{(\mathcal{S}, \mathcal{T}, \sigma_{AB})} P^{\text{game}}(\mathcal{S}, \mathcal{T}, \sigma_{AB}) \quad (2.59)$$

2.6.3 Steering

Steering determines the states that Alice can remotely prepare on Bob's subsystem by making local measurements of her subsystem. If Alice and Bob share a quantum state σ_{AB}, Bob's part of the shared state is $\sigma_B = \mathrm{Tr}_A \sigma_{AB}$. In a theory with convex state space $\mathcal{C}$ (of which quantum theory is an example), one can decompose Bob's part of the shared state, σ_B, in many different ways as a convex sum:

$$\sigma_B = \sum_a p(a|s)\sigma_{s,a} \tag{2.60}$$

where $\sigma_{s,a} \in \mathcal{C}$ is the state of Bob's subsystem if Alice gets outcome a for measurement s. This decomposition corresponds to the ensemble $\mathcal{E}_s = \{p(a|s), \sigma_{s,a}\}_a$. By making measurement s Alice thus steers Bob's subsystem to the ensemble $\mathcal{E}_s$. Steering is constrained by the no-signalling condition, i.e.,

$$\sigma_B = \sum_a p(a|s)\sigma_{s,a} = \sum_a p(a|s')\sigma_{s',a}, \quad \forall \quad s, s' \in \mathcal{S} \tag{2.61}$$

For any set of ensembles $\{\mathcal{E}_s\}_s$ that satisfy no-signalling (2.61), and therefore consistently define a σ_B, one can find a bipartite quantum state σ_{AB} and measurements that allow Alice to steer to such ensembles.

2.6.4 Result

The main result of this paper [19] is that the fine-grained uncertainty relations (fgURs) for Bob's measurements on the steerable states of his subsystem limit the strength of nonlocality in any theory. In quantum theory, for all games where Bob's optimal measurements are known, the set of steerable states is the same as the set of maximally certain states on Bob's subsystem. As we will see below, the fgURs for Bob's measurements restrict the winning probability in a nonlocal game. We sketch the results for the case of XOR games and refer the reader to [19] for further details and additional results regarding uncertainty relations and their link with nonlocality.

$$
\begin{aligned}
P^{\text{game}}(\mathcal{S},\mathcal{T},\sigma_{AB}) &= \sum_{s,t} p(s,t) \sum_a p(a, b = x^{(t)}_{s,a} | s,t)_{\sigma_{AB}} &(2.62)\\
&= \sum_s p(s) \sum_t p(t) \sum_a p(a|s) p(b = x^{(t)}_{s,a}|t)_{\sigma_{s,a}} &(2.63)\\
&= \sum_s p(s) \sum_a p(a|s) \sum_t p(t) p(b = x^{(t)}_{s,a}|t)_{\sigma_{s,a}} &(2.64)\\
&\leq \sum_s p(s) \sum_a p(a|s) \zeta_{\vec{x}_{s,a}}(\mathcal{T},\mathcal{D}) &(2.65)\\
&\quad (\because \zeta_{\vec{x}_{s,a}}(\mathcal{T},\mathcal{D}) \equiv \max_{\sigma_{s,a}} \sum_t p(t) p(x^{(t)}_{s,a}|t)_{\sigma_{s,a}})\\
&\leq \sum_s p(s) \sum_a p(a|s) \max_{\vec{x}_{s,a}} \zeta_{\vec{x}_{s,a}}(\mathcal{T},\mathcal{D}) &(2.66)\\
&= (\sum_s p(s) \sum_a p(a|s)) \zeta_{\vec{x}}(\mathcal{T},\mathcal{D}) &(2.67)\\
&= \zeta_{\vec{x}}(\mathcal{T},\mathcal{D}) &(2.68)
\end{aligned}
$$

Clearly, $P^{\text{game}}(\mathcal{S},\mathcal{T},\sigma_{AB}) \leq \zeta_{\vec{x}}(\mathcal{T},\mathcal{D})$ establishes the link between nonlocality and fgURs. The winning probability is restricted by uncertainty relations for Bob's measurement settings. Now, an optimal strategy $(\mathcal{S},\mathcal{T}_{opt},\sigma_{AB})$ would involve a choice of measurement settings for Bob, $\mathcal{T}_{opt}$, that maximizes the winning probability in the game. Also, we need to look at the states that Alice can prepare on Bob's subsystem (the "steerable states") given her measurement settings $\mathcal{S}$ rather than all possible quantum states that Bob's subystem could be in if we only imposed positive semidefiniteness, and unit trace, of the density matrix. In (2.65) the maximization is over all possible $\sigma_{s,a}$ that satisfy the condition to be a valid quantum state. In the context of the game, though, one needs to maximize over a subset of these quantum states – the set of steerable states for measurement settings $\mathcal{S}$, i.e.,

$$
P^{\text{game}}_{\max} = \max_{\{\mathcal{E}_s\}_s} \sum_s p(s) \sum_a p(a|s) \zeta^{\sigma_{s,a}}_{\vec{x}}(\mathcal{T}_{opt},\mathcal{D}) \qquad (2.69)
$$

which follows from

$$
\begin{aligned}
P^{\text{game}}(\mathcal{S},\mathcal{T},\sigma_{AB}) &= \sum_s p(s)\sum_a p(a|s)\sum_t p(t)p(b=x^{(t)}_{s,a}|t)_{\sigma_{s,a}} && (2.70)\\
&\leq \sum_s p(s)\sum_a p(a|s)\zeta^{\sigma_{s,a}}_{\vec{x}_{s,a}}(\mathcal{T},\mathcal{D}) && (2.71)\\
&\quad (\zeta^{\sigma_{s,a}}_{\vec{x}_{s,a}}(\mathcal{T},\mathcal{D}) \equiv \max_{\sigma_{s,a}\in\mathcal{E}_s}\sum_t p(t)p(x^{(t)}_{s,a}|t)_{\sigma_{s,a}}) && (2.72)\\
&\leq \sum_s p(s)\sum_a p(a|s)\zeta^{\sigma_{s,a}}_{\vec{x}}(\mathcal{T},\mathcal{D}) && (2.73)\\
&\quad (\zeta^{\sigma_{s,a}}_{\vec{x}}(\mathcal{T},\mathcal{D}) \equiv \max_{\vec{x}_{s,a}}\zeta^{\sigma_{s,a}}_{\vec{x}_{s,a}}(\mathcal{T},\mathcal{D})) && (2.74)\\
&\leq \sum_s p(s)\sum_a p(a|s)\zeta^{\sigma_{s,a}}_{\vec{x}}(\mathcal{T}_{opt},\mathcal{D}) && (2.75)\\
&\leq \max_{\{\mathcal{E}_s\}_s}\sum_s p(s)\sum_a p(a|s)\zeta^{\sigma_{s,a}}_{\vec{x}}(\mathcal{T}_{opt},\mathcal{D}) && (2.76)
\end{aligned}
$$

Clearly,

$$
\begin{aligned}
P^{\text{game}}_{\max} &= \max_{\{\mathcal{E}_s\}_s}\sum_s p(s)\sum_a p(a|s)\zeta^{\sigma_{s,a}}_{\vec{x}}(\mathcal{T}_{opt},\mathcal{D})\\
&\leq \sum_s p(s)\sum_a p(a|s)\zeta_{\vec{x}}(\mathcal{T}_{opt},\mathcal{D}) && (2.77)
\end{aligned}
$$

In general, the maximally certain state for $\mathcal{T}_{opt}$, $\rho_{\vec{x}}$, that achieves the upper bound in (2.77) may not be steerable, i.e., Alice may not be able to prepare that state given her measurement settings $\mathcal{S}$. The question, therefore, is: Does $\rho_{\vec{x}} \in \{\mathcal{E}_s\}_s$? If it does, then the winning probability in the game is the maximum it can be without violating the fine-grained uncertainty relations, i.e., the upper bound in (2.77) can be realized. If it doesn't, then the upper bound cannot be realized because of limited steering. If one wants to do better than the upper bound in (2.77) one would necessarily have to violate the uncertainty principle embodied in the fgURs. It turns out that in quantum mechanics (at least for all XOR games) $\rho_{\vec{x}} \in \{\mathcal{E}_s\}_s$ and the fgURs for Bob's measurements $\mathcal{T}_{opt}$ give a tight bound:

$$
P^{\text{game}}_{\max} = \sum_s p(s)\sum_a p(a|s)\zeta_{\vec{x}}(\mathcal{T}_{opt},\mathcal{D})
$$

Whether this holds for all games in quantum theory is still an open question. The general insight here is that limited steering and strong uncertainty relations restrict nonlocality and the winning probability in a nonlocal game depends on the tradeoff between steering and fgURs – whether the maximally certain states of the fgURs are also steerable. We sketch the proof, adapted from [19], for XOR games in quantum theory:

Proof. We consider here the Bell scenario $(2, M, 2)$, i.e., two parties with dichotomic measurements. We denote the measurement outcomes for Alice and Bob by $a \in \{0,1\}$ and $b \in \{0,1\}$ respectively. By an XOR game we refer to games where the winning correlations satisfy some condition on the XOR of the outputs, i.e., $a \oplus b$ should satisfy some constraint. One can represent the winning condition as a predicate: $V(c = a \oplus b|s,t) = 1$ if and only $a \oplus b = c$ are the winning answers, given s and t. Otherwise, $V(c = a \oplus b|s,t) = 0$.

Structure of measurements: We denote the measurement operators for Alice and Bob by A_s^a and B_t^b respectively, satisfying $A_s^a A_s^{a'} = \delta_{a,a'} A_s^a$ and $B_t^b B_t^{b'} = \delta_{b,b'} B_t^b$. We can write the observables as:

$$A_s = A_s^0 - A_s^1 \tag{2.78}$$

$$B_t = B_t^0 - B_t^1 \tag{2.79}$$

with eigenvalues ± 1, where we label outcome '$+1$' as '0' and '-1' as '1'. Tsirelson showed [7, 31] that the optimal measurements $(\mathcal{S}, \mathcal{T})$ that yield the maximum winning probability in quantum theory are traceless observables of the form

$$A_s = \sum_j a_s^{(j)} \Gamma_j \equiv \vec{a}_s . \vec{\Gamma} \tag{2.80}$$

$$B_t = \sum_j b_t^{(j)} \Gamma_j \equiv \vec{b}_t . \vec{\Gamma} \tag{2.81}$$

where $\vec{a}_s = (a_s^{(1)}, ..., a_s^{(N)}) \in \mathbb{R}^N$ and $\vec{b}_t = (b_t^{(1)}, ..., b_t^{(N)}) \in \mathbb{R}^N$ are real unit vectors, and $N = \min(|\mathcal{S}|, |\mathcal{T}|)$. $\Gamma_1, ..., \Gamma_N$ are the anticommuting generators of a Clifford algebra:

$$\{\Gamma_j, \Gamma_k\} = 2\delta_{jk}\mathbb{I}, \quad \forall j,k \in \{1,2,...,N\} \tag{2.82}$$

Using the fact that $A_s^{(0)} + A_s^{(1)} = \mathbb{I}$ and $B_t^{(0)} + B_t^{(1)} = \mathbb{I}$ we can write the measurement operators as:

$$A_s^a = \frac{1}{2}(\mathbb{I} + (-1)^a A_s) \tag{2.83}$$

$$B_t^b = \frac{1}{2}(\mathbb{I} + (-1)^b B_t) \tag{2.84}$$

Optimal state: Tsirelson also showed that the optimal state that Alice and Bob share, corresponding to the optimal measurements above, is the maximally entangled state:

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} |k\rangle|k\rangle \tag{2.85}$$

where $d = (2^{\lfloor N/2 \rfloor})^2$. Obviously, $\sigma_{AB} = |\Psi\rangle\langle\Psi| = \frac{1}{d}\sum_{k,l} |k,k\rangle\langle l,l|$. The winning probability for this XOR game is

$$\begin{aligned}
P^{\text{game}}(\mathcal{S},\mathcal{T},\sigma_{AB}) &= \sum_{s,t} p(s,t) \sum_{a,b} V(c = a \oplus b|s,t) \langle\Psi|A_s^a \otimes B_t^b|\Psi\rangle \\
&= \sum_{s,t} p(s,t) \sum_{a,b} V(c = a \oplus b|s,t) \\
&\quad \langle\Psi|\frac{1}{2}(\mathbb{I} + (-1)^a A_s) \otimes \frac{1}{2}(\mathbb{I} + (-1)^b B_t)|\Psi\rangle \\
&= \frac{1}{2}\sum_{s,t} p(s,t) \sum_{c} V(c|s,t)(1 + (-1)^c \langle\Psi|A_s \otimes B_t|\Psi\rangle)
\end{aligned}$$

Therefore,

$$P^{\text{game}}(\mathcal{S},\mathcal{T},\sigma_{AB}) = \frac{1}{2}\sum_{s,t} p(s,t) \sum_{c} V(c|s,t)(1+(-1)^c \langle\Psi|A_s \otimes B_t|\Psi\rangle) \tag{2.86}$$

Steering to maximally certain states: Now that we know the optimal measurements for Bob, we can write down the corresponding uncertainty operators. Given these uncertainty operators we will show that Alice can steer Bob's state to the maximally certain states for these operators. Each element in the set of Bob's measurements $\{B_t\}_{t\in\mathcal{T}}$ is decomposable in terms of POVM elements $\{B_t^b\}_{b\in\{0,1\}}$ and the $\zeta_{\vec{x}}$ thus correspond to the largest eigenvalue of the operator

$$Q_{\vec{x}_{s,a}} = \sum_t p(t) B_t^{x_{s,a}^{(t)}} \tag{2.87}$$

For XOR games,

$$
\begin{aligned}
Q_{\vec{x}_{s,a}} &= \sum_t p(t) \sum_{b,V(a,b|s,t)=1} B_t^b & (2.88)\\
&= \frac{1}{2}\sum_t p(t) \sum_{b,V(a,b|s,t)=1} (\mathbb{I} + (-1)^b B_t) & (2.89)\\
&= \frac{1}{2}\sum_t p(t)(|\sum_b V(a,b|s,t)|\mathbb{I} + \sum_{b,V(a,b|s,t)=1} (-1)^b \vec{b}_t.\vec{\Gamma}) & (2.90)\\
&= \frac{1}{2}(c_{s,a}\mathbb{I} + \vec{v}_{s,a}.\vec{\Gamma}) & (2.91)
\end{aligned}
$$

where

$$
\begin{aligned}
c_{s,a} &\equiv \sum_t p(t)|\sum_b V(a,b|s,t)| & (2.92)\\
\vec{v}_{s,a} &\equiv \sum_t p(t) \sum_{b,V(a,b|s,t)=1} (-1)^b \vec{b}_t & (2.93)
\end{aligned}
$$

Maximally certain state: For any XOR game and for any uncertainty operator $Q_{\vec{x}_{s,a}}$,

$$
\begin{aligned}
\zeta_{\vec{x}_{s,a}} &= \max_{\rho\geq 0, \mathrm{Tr}\rho=1} \mathrm{Tr}(\rho Q_{\vec{x}_{s,a}}) & (2.94)\\
&= \mathrm{Tr}(\rho_{\vec{x}_{s,a}} Q_{\vec{x}_{s,a}}) & (2.95)\\
&= \frac{1}{2}(c_{s,a} + ||\vec{v}_{s,a}||_2) & (2.96)\\
& & (2.97)
\end{aligned}
$$

where the maximally certain state is

$$
\rho_{\vec{x}_{s,a}} = \frac{1}{d}(\mathbb{I} + \sum_j r_{s,a}^{(j)} \Gamma_j) \qquad (2.98)
$$

and

$$
\vec{r}_{s,a} = \vec{v}_{s,a}/||\vec{v}_{s,a}||_2 \qquad (2.99)
$$

Of course,

$$
||\vec{v}_{s,a}||_2 = (\sum_j (v_{s,a}^{(j)})^2)^{\frac{1}{2}}
$$

We refer to [19] for the proof. We need to show that the ensemble of maximally certain states can be steered to by Alice. This is indeed the case in quantum theory:

Steerable states: For any XOR game we have $\forall s, \hat{s} \in \mathcal{S}$,

$$\sum_a p(a|s)\rho_{\vec{x}_{s,a}} = \sum_a p(a|\hat{s})\rho_{\hat{s},a} \tag{2.100}$$

That is, any ensemble of maximally certain states is steerable. See [19] for the proof.

Tsirelson's bound: One can recover Tsirelson's bound using the maximally certain states $\rho_{\vec{x}_{s,a}}$ and Bob's corresponding optimal measurement operators $B_t^{x_{s,a}^{(t)}}$. The maximum winning probability in the CHSH game is given by:

$$\begin{aligned}
P_{\max}^{\text{game}} &= \frac{1}{4}\sum_{s,a\in\{0,1\}} \mathrm{Tr}(\rho_{\vec{x}_{s,a}} Q_{\vec{x}_{s,a}}) && (2.101)\\
&= \frac{1}{4}\sum_{s,a,t\in\{0,1\}} p(t)\mathrm{Tr}(\rho_{\vec{x}_{s,a}} B_t^{x_{s,a}^{(t)}}) && (2.102)\\
&= \frac{1}{4}\sum_{s,a,t\in\{0,1\}} \frac{1}{2}\left(\frac{1}{2}+\frac{1}{2\sqrt{2}}\right) && (2.103)\\
&= \left(\frac{1}{2}+\frac{1}{2\sqrt{2}}\right) && (2.104)
\end{aligned}$$

which is Tsirelson's bound. Here we have used the fact that:

$$\begin{aligned}
\mathrm{Tr}(\rho_{\vec{x}_{s,a}} B_t^{x_{s,a}^{(t)}}) &= \frac{1}{2}\mathrm{Tr}\left(\rho_{\vec{x}_{s,a}}(\mathbb{I} + (-1)^{x_{s,a}^{(t)}} B_t)\right) && (2.105)\\
&= \frac{1}{2}\left(1 + (-1)^{x_{s,a}^{(t)}}\mathrm{Tr}(\rho_{\vec{x}_{s,a}} B_t)\right) && (2.106)\\
&= \frac{1}{2}(1 + (-1)^{x_{s,a}^{(t)}}\vec{r}_{s,a}.\vec{b}_t) && (2.107)\\
&= \frac{1}{2}\left(1 + (-1)^{x_{s,a}^{(t)}}\frac{(-1)^{x_{s,a}^{(t)}}}{\sqrt{2}}\right) && (2.108)\\
&= \frac{1}{2}\left(1+\frac{1}{\sqrt{2}}\right) && (2.109)
\end{aligned}$$

2.7 Chapter Summary

We have reviewed progress towards finding informational principles that could potentially characterize quantum correlations. We have also reviewed the connection between the uncertainty principle and nonlocality, pointed out by Oppenheim and Wehner [19]. The problem of characterizing quantum correlations is, of course, not completely settled by any of these approaches and despite partial answers the question remains wide open, especially in the multipartite case, $N > 2$.

Chapter 3

Future Directions

It would be interesting to consider the efficacy of a principle like information causality (IC) in recovering the quantum bound in nonlocality-without-inequality type arguments, e.g., those proposed by Hardy and Cabello, especially in the multipartite scenario. In the two-qubit case, it has been shown [25] that under the principle of information causality imposed on no-signalling correlations the quantum bound in either case (Hardy's argument, and Cabello's) is not recovered. In fact, while IC takes one closer to the quantum bound than no-signalling alone, the bound given by IC is the same for both scenarios. On the other hand, the quantum bounds are different for the two scenarios. Recovering the quantum bound in these scenarios, while also obtaining Tsirelson's bound, by some modification of IC could be explored. A consideration of Cabello's argument for three parties, under a suitable generalization of IC to this scenario, could offer clues regarding the power or limitation of IC in this multipartite scenario. Further, it has been shown [30] that there exist no-signalling TOBL (time-ordered-bilocal) correlations outside the quantum set that satisfy any bipartite principle like information causality. Excluding superquantum TOBL correlations is another constraint that any candidate physical principle should satisfy.

We briefly review the limitations pointed out in [25, 30].

3.1 Limitations of proposed physical principles

3.1.1 Hardy's nonlocality argument vis-à-vis Information Causality

We review the result of [25] demonstrating a limitation of information causality. Consider the case of two spacelike separated parties, Alice and Bob, each with a qubit on which (s)he can perform any one of two measurements – A_0 and A_1 for Alice, and B_0 and B_1 for Bob. Each measurement has two possible outcomes $\{0,1\}$. Alice's and Bob's measurements, A_x and B_y, are respectively labelled by $x, y \in \{0,1\}$, and their outcomes, $(-1)^a$ and $(-1)^b$, by $a, b \in \{0,1\}$. The joint probabilities of these measurements, $p(a,b|x,y)$, are constrained by the following equations (apart from the usual positivity, normalization, and no-signalling constraints):

$$\begin{aligned} p(0,0|0,0) &= q_0 \\ p(1,1|0,1) &= 0 \\ p(1,1|1,0) &= 0 \\ p(1,1|1,1) &= q_3 \end{aligned} \tag{3.1}$$

These equations contradict local realism (and therefore rule out local realistic models) if $q_0 < q_3$. This becomes clear if one considers the case when a local realistic model assigns the values $A_1 = -1, B_1 = -1$ to the respective measurements of Alice and Bob. This immediately implies that $B_0 = +1$ and $A_0 = +1$. Clearly, then, $q_0 \geq q_3$ for local realism to hold. $q_0 < q_3$, therefore, implies a violation of local realism. The special case, $q_0 = 0$, is called Hardy's paradox [32, 33], while the general case was obtained by Cabello [34]. The eight-dimensional no-signalling polytope in the $(2,2,2)$ scenario consists of 24 vertices, 16 local and 8 nonlocal. The 16 vertices of the local polytope are given by

$$\begin{aligned} p^{\alpha\beta\gamma\delta}_{ab|xy} &= 1, \text{ if } a = \alpha x \oplus \beta, b = \gamma y \oplus \delta; \\ &= 0, \text{ otherwise,} \end{aligned} \tag{3.2}$$

where $\alpha, \beta, \gamma, \delta \in \{0,1\}$. The 8 nonlocal vertices are given by

$$\begin{aligned} p^{\alpha\beta\gamma}_{ab|xy} &= \frac{1}{2}, \text{ if } a \oplus b = xy \oplus \alpha x \oplus \beta y \oplus \gamma; \\ &= 0, \text{ otherwise,} \end{aligned} \tag{3.3}$$

where $\alpha, \beta, \gamma \in \{0,1\}$. Five local vertices and one nonlocal vertex satisfy equations (3.1): $p^{0001}_{ab|xy}$, $p^{0011}_{ab|xy}$, $p^{0100}_{ab|xy}$, $p^{1100}_{ab|xy}$, $p^{1111}_{ab|xy}$, and $p^{001}_{ab|xy}$. The joint probabilities consistent with Hardy's conditions (3.1), when $q_0 = 0$, are given by the convex sum of these vertices:

$$p^{\mathcal{H}}_{ab|xy} = c_1 p^{0001}_{ab|xy} + c_2 p^{0011}_{ab|xy} + c_3 p^{0100}_{ab|xy} + c_4 p^{1100}_{ab|xy} + c_5 p^{1111}_{ab|xy} + c_6 p^{001}_{ab|xy} \quad (3.4)$$

where $\sum_{i=1}^{6} c_i = 1$. Allowing $q_0 \neq 0$ ($q_0 < q_3$), equations (3.1) admit another four local vertices and one nonlocal vertex: $p^{0000}_{ab|xy}$, $p^{0010}_{ab|xy}$, $p^{1000}_{ab|xy}$,$p^{1010}_{ab|xy}$, and $p^{110}_{ab|xy}$. The joint probabilities obeying these conditions are of the form:

$$p^{\mathcal{C}}_{ab|xy} = p^{\mathcal{H}}_{ab|xy} + c_7 p^{0000}_{ab|xy} + c_8 p^{0010}_{ab|xy} + c_9 p^{1000}_{ab|xy} + c_{10} p^{1010}_{ab|xy} + c_{11} p^{110}_{ab|xy} \quad (3.5)$$

where $\sum_{i=1}^{11} c_i = 1$. Clearly, the success probability for Hardy's argument is given by $p^{\mathcal{H}}_{11|11} = \frac{1}{2}c_6$. The maximum, $(p^{\mathcal{H}}_{11|11})_{max} = \frac{1}{2}$, is achieved for $c_6 = 1$, and $c_1 = c_2 = c_3 = c_4 = c_5 = 0$. The success probability for Cabello's argument is $p^{\mathcal{C}}_{11|11} - p^{\mathcal{C}}_{00|00} = (\frac{1}{2}c_6 + c_{10}) - (c_7 + c_8 + c_9 + c_{10} + \frac{1}{2}c_{11})$. $(p^{\mathcal{C}}_{11|11} - p^{\mathcal{C}}_{00|00})_{max} = \frac{1}{2}$ for $c_6 = 1$, and the other $c_i's = 0$. No-signalling alone, therefore, restricts the success probability to 0.5 in both scenarios. On applying the restriction of information causality, on the other hand, the authors of [25] obtain a maximum success probability of 0.20717 in both cases. The maximum success probability in quantum theory is 0.09 for Hardy's scenario and 0.1078 for Cabello's argument. This discrepancy hints at a need for either refining information causality further or searching for an alternative principle that recovers bounds on quantum correlations in a more comprehensive sense.

3.1.2 The need for multipartite principles

Characterizing the set of quantum correlations is still very much an open problem. Principles like non-trivial communication complexity [9], information causality [8] and macroscopic locality [18] – while providing partial insights into quantum nonlocality – have not been shown to completely characterize the set of quantum correlations yet. Indeed, there is some recent evidence [30] suggesting that quantum correlations require intrinsically multipartite principles. We review the result of [30], demonstrating the existence of stronger-than-quantum correlations satisfying any bipartite information principle (like no-signalling, information causality). Such superquantum correlations are not ruled out by information causality and thus demonstrate a limitation of this approach.

TOBL correlations: Consider tripartite correlations $P(a_1a_2a_3|x_1x_2x_3)$ that are local with respect to any bipartition, i.e., they satisfy Bell-locality, and therefore no-signalling (and in general, any bipartite information principle) for all possible bipartitions of the three parties. $P(a_1a_2a_3|x_1x_2x_3)$ admits a *time-ordered bi-local* (TOBL) model if it can be expressed as

$$\begin{aligned} P(a_1a_2a_3|x_1x_2x_3) &= \sum_\lambda p_\lambda^{i|jk} P(a_i|x_i,\lambda) P_{j\to k}(a_ja_k|x_jx_k,\lambda) \\ &= \sum_\lambda p_\lambda^{i|jk} P(a_i|x_i,\lambda) P_{j\leftarrow k}(a_ja_k|x_jx_k,\lambda) \end{aligned} \tag{3.6}$$

where $(i,j,k)=(1,2,3),(2,3,1),(3,1,2)$, and $P_{j\to k}$, $P_{j\leftarrow k}$ satisfy:

$$P_{j\to k}(a_j|x_j,\lambda) = \sum_{a_k} P_{j\to k}(a_ja_k|x_jx_k,\lambda) \tag{3.7}$$

$$P_{j\leftarrow k}(a_k|x_k,\lambda) = \sum_{a_j} P_{j\leftarrow k}(a_ja_k|x_jx_k,\lambda) \tag{3.8}$$

The operational meaning here is that under a bipartition $i|jk$, parties j and k are allowed to collaborate (via "wirings" – using the output of one as input for the other). Depending on which way the one-way communication happens (either $j\to k$ or $j\leftarrow k$), the outcomes obey the corresponding statistics. Since the TOBL model is local under any bipartition it will satisfy any bipartite principle satisfied by quantum correlations. It turns out (cf. [30]) that for any general protocol, given NS-boxes producing TOBL correlations, the final probability distribution belongs to the set of TOBL correlations.

GYNI inequality: We consider the GYNI (Guess-Your-Neighbour's-Input) inequality (cf. [29]) that gives the same bound for both classical and quantum correlations in the Bell scenario $(3,2,2)$, i.e., three parties with two dichotomic measurements each:

$$\begin{aligned} \mathbf{B}(P_{\mathrm{Q}}) =& P_{\mathrm{Q}}(000|000) + P_{\mathrm{Q}}(110|011) \\ &+ P_{\mathrm{Q}}(011|101) + P_{\mathrm{Q}}(101|110) \leq 1. \end{aligned} \tag{3.9}$$

If one maximizes this expression (3.9) over the set of $(3,2,2)$ TOBL correlations, one obtains a value of $B_{max}=\frac{7}{6}$, thus violating the GYNI inequality. The TOBL probability distribution attaining this bound can be found in the Supplemental Material of [30]. What concerns us here is the fact that this superquantum TOBL correlation satisfies any bipartite information principle by virtue of belonging to the TOBL set. Principles like information causality [8] cannot exclude such correlations and are therefore of limited value in their bipartite incarnation. This suggests that any information-theoretic principle seeking to characterize quantum correlations will have to be intrinsically multipartite.

3.2 Conclusion

The limitations of physical principles in characterizing the set of quantum correlations pointed out in this chapter raise questions that need to be answered if this endeavour is to progress any further. Indeed, their limitations offer potential clues towards modifying or generalizing these physical principles to recover quantum correlations from them.

Bibliography

[1] J. S. Bell, Physics **1,** 195 (1964).

[2] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2nd edition, Cambridge University Press (2004).

[3] G.M. Ziegler, *Lectures on polytopes*, Vol. 152, Springer (1995)

[4] B. Grünbaum, *Convex Polytopes*, 2nd edition, Graduate Texts in Mathematics, **221**, Springer-Verlag (Berlin) (2003).

[5] Matty J. Hoban, *Computational Perspectives on Bell Inequalities and Many-body Quantum Correlations*, PhD thesis, quant-ph/1204.5913

[6] F. Clauser, M. A. Horne, A. Shimony and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[7] B. Tsirelson. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4:93–100, 1980.

[8] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, M. Żukowski, *Information causality as a physical principle*, Nature **461**, 1101 (2009)

[9] W. van Dam, *Implausible Consequences of Superstrong Nonlocality*, quant-ph/0501159.

[10] Yeong-Cherng Liang, *Correlations, Bell Inequality Violation and Quantum Entanglement*, PhD thesis, quant-ph/0810.5400

[11] Nicolas Gisin, *Bell inequalities: many questions, a few answers*, arXiv:quant-ph/0702021v2

[12] *All the Bell Inequalities*, `http://qig.itp.uni-hannover.de/qiproblems/1`

[13] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. 49, 1804 (1982); G. Weihs et al., Phys. Rev. Lett. 81, 5039 (1998); M. A. Rowe et al., Nature (London) 409, 791 (2001).

[14] S. Popescu and D. Rohrlich, *Quantum nonlocality as an axiom*, Found. Phys. **24**, 379 (1994).

[15] Nielsen, M.A. and Chuang, I.L., *Quantum information and computation*, Cambridge University Press, (2000)

[16] Marcin Pawłowski, Valerio Scarani, *Information Causality*, quant-ph/1112.1142

[17] Masanes, Ll. and Acin, A. and Gisin, N., Phys. Rev. A 73, 012112 (2006)

[18] M. Navascués and H. Wunderlich, *A glance beyond the quantum model*, Proc. Roy. Soc. Lond. A **466**, 881-890 (2009)

[19] J. Oppenheim, and S. Wehner, *The Uncertainty Principle Determines the Nonlocality of Quantum Mechanics*, Science **330**, 6007, 1072-1074 (2010).

[20] Valerio Scarani, *Feats, Features and Failures of the PR-box*, quant-ph/0603017

[21] N. Cerf, N. Gisin, S. Massar, S. Popescu, *Phys. Rev. Lett.* **94**, 220403 (2005)

[22] B. F. Toner, D. Bacon, *Phys. Rev. Lett.* **91**, 187904 (2003).

[23] J. Barrett, Phys. Rev. A 75, 032304 (2007).

[24] D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten, Phys. Rev. Lett. 104, 080402 (2010)

[25] Ahanj, A. and Kunkri, S. and Rai, A. and Rahaman, R. and Joag, P.S. *Bound on Hardy's nonlocality from the principle of information causality*, Phys. Rev. A 81, 032103 (2010)

[26] Wehner, S. and Winter, A., *Entropic uncertainty relations—a survey*, New Journal of Physics, 12, 025009, (2010)

[27] Navascués, M. and Pironio, S. and Acín, A. *A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations*, New Journal of Physics, **10**, 073013 (2008)

[28] Cavalcanti, D. and Salles, A. and Scarani, V. *Macroscopically local correlations can violate information causality*, Nature Communications, 1, 136, (2010)

[29] M.L. Almeida, J.D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, Phys. Rev. Lett. 104, 230404 (2010)

[30] R. Gallego, L.E. Würflinger, A. Acín, and M. Navascués, *Quantum correlations require multipartite information principles*, Phys. Rev. Lett. 107, 210403 (2011)

[31] B. Tsirelson. Quantum analogues of Bell inequalities: The case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557–570, 1987.

[32] L. Hardy, Phys. Rev. Lett. 68, 2981 (1992).

[33] L. Hardy, Phys. Rev. Lett. 71, 1665 (1993).

[34] A. Cabello, Phys. Rev. A 65, 032108 (2002).