Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

Una perspectiva teórico-computacional sobre fundamentos de la información cuántica

Tesis presentada para optar al título de
Doctor de la Universidad de Buenos Aires en el área Ciencias de la Computación

Lic. Gabriel Ignacio Senno

Directores de tesis:	Dr. Ariel Martín Bendersky
	Dr. Santiago Daniel Figueira
Consejero de estudios:	Dr. Diego Garbervetsky

Lugar de trabajo: Instituto de Investigación en Ciencias de la Computación (ICC).

Buenos Aires, Abril 2017.

Fecha de defensa: 27/04/17.

UNA PERSPECTIVA TEÓRICO-COMPUTACIONAL SOBRE FUNDAMENTOS DE LA INFORMACIÓN CUÁNTICA

La presente tesis contiene resultados sobre fundamentos de la teoría cuántica de la información obtenidos mediante conexiones novedosas con las teorías de la computabilidad y la complejidad comunicacional.

En la primera parte, presentamos consecuencias de, como es usual en los experimentos, usar pseudoaleatoriedad en lugares donde la teoría cuántica asume aleatoriedad. Obtenemos tres resultados:

El primero consiste en un nuevo *loophole* para experimentos de Bell. Probamos, usando herramientas de la teoría de la inferencia inductiva, que elegir las entradas en un experimento de Bell usando generadores de números pseudoaleatorios permite a un adversario, bajo ciertas asunciones razonables, preparar de manera local cajas que dan lugar a una estadística no-local.

En segundo lugar, damos un protocolo que permite, dadas cajas no-locales que generan sus salidas de manera computable y con ayuda de algún mecanismo posiblemente escondido de señalización, extraer tal mecanismo para su uso como canal de comunicación, con el sólo conocimiento de una cota a la complejidad computacional de las cajas.

El tercer y último aporte de esta primera parte consiste en un protocolo que permite distinguir, a través del uso de tests de Martin-Löf, cualquier mezcla pseudoaleatoria de estados cuánticos del estado máximamente mixto. Se incluyen también los resultados de una realización experimental de un caso especial del protocolo llevada a cabo por el grupo del Dr. Miguel Larotonda.

En la segunda parte, retomamos el estudio de la no-localidad de Bell pero esta vez desde una perspectiva informacional. Más precisamente, investigamos la relación entre la ventaja que la cuántica ofrece en el modelo de complejidad comunicacional de funciones, y su carácter no-local. Una de las técnicas más ajustadas para probar cotas inferiores a la complejidad comunicacional clásica se conoce como *partition-bound*. El resultado principal de esta segunda parte consiste en dar un método para extraer grandes violaciones de desigualdades de Bell de todo protocolo cuántico que compute una dada función comunicando menos qbits que su valor de partition-bound asociado. Ésto aplica a la mayoría de las funciones usualmente estudiadas en complejidad comunicacional. Las violaciones que obtenemos son resistentes al *loophole de la detección* y mostramos como también pueden hacerse resistentes a ruido uniforme.

Palabras claves: fundamentos de la mecánica cuántica, no-localidad de Bell, pseudoaleatoriedad, aleatoriedad algorítmica, preparación de estados mixtos, loophole de Bell, complejidad comunicacional cuántica.

A COMPUTER-THEORETICAL OUTLOOK ON FOUNDATIONS OF QUANTUM INFORMATION

The present thesis contains results on foundations of quantum information theory obtained through new connections with computability theory and communication complexity.

In the first part, we give consequences of using, as is usual done in experiments, pseudorandomness where quantum theory assumes randomness. We obtain three results:

First, we present a new loophole for Bell-like experiments. We prove, using tools from the theory of inductive inference, that choosing the inputs for a Bell tests using private pseudorandom number generators allows an adversary, under reasonable assumptions, to predict forthcoming inputs and prepare local boxes that seem non-local.

Second, we give a protocol that, given non-local boxes generating their outputs computably and with the aid of a (possibly hidden) signaling mechanism, extract such mechanism and turn it into a communication channel, provided a bound on the computational complexity of the boxes is known. We arrive to this result by proving a novel connection between the theories of inductive inference and algorithmic randomness.

Third, through the use of Martin-Löf tests, we give a protocol to distinguish any pseudorandom mixture of quantum states from the maximally mixed state. Furthermore, a proof-of-concept experiment for an special case of this protocol done by the group of Dr. Miguel Larotonda is presented.

In the second part, we come back to Bell non-locality, but with an informational perspective. We concentrate on the relationship between the advantage that quantum mechanics offers in communication complexity and its non-local nature. One of the strongest techniques known to prove lower bounds on classical communication complexity is the so-called partition bound. We show how to derive large Bell violations from any problem whose quantum communication complexity is smaller than its partition bound value. This applies to most of the functions usually studied in communication complexity theory. The violations we get are resistant to the detection loophole, can be exponential in the size of the inputs and we show that they can also be made resistant to uniform noise.

Keywords: quantum foundations, Bell non-locality, pseudorandomness, algorithmic randomness, mixed states preparation, Bell loophole, quantum communication complexity.

CONTENTS

We, the authors. The present thesis is written in first-person plural. Since different parts of the work were done in collaboration with different sets of people, different instances of we may mean different sets.

1. INTRODUCTION AND PRELIMINARES

1.1 Background and motivation

Quantum computation and quantum information theory [NC11] are about using the framework of quantum mechanics to design protocols and/or build systems that perform information-processing tasks which are classically difficult or even impossible to achieve. This application of quantum mechanical principles to the study of information-processing task has provided remarkable advances such as: Shor's quantum algorithm to efficiently factor integers [Sho97], key distribution protocols which base their security on the laws of physics instead of the hardness of some mathematical task [BB84, Eke91] and full randomness amplification and expansion protocols [CK11, CR12, GMDLT$^+$13], to name a few.

The aim of this thesis is to analyze some of the fundamental quantum mechanical concepts, key to the aforementioned results, from a computer-theoretical perspective. The approach will be two-fold: on the one hand, we will analyse assumptions behind the results; on the other, we will trust the assumptions and then quantify the computational gain of applying those quantum mechanical features. Specifically, on the assumptions side, we will restort to the theories of algorithmic randomness and inductive inference to study the impact of replacing the assumption of randomness by the use of pseudorandomness in some areas of the theory. Then, on the applications side, we will look at the advantage that quantum mechanics offers in the area of communication complexity. In the next paragraphs we outline the main questions we will be dealing with.

1.1.1 Randomness vs pseudorandomness in non-locality experiments

The fact that measuring a property of a quantum system can instantaneously determine the results of another property measured on a distant system is on of the features of quantum mechanics that has puzzled scientists the most since its origin. Indeed, such kind of non-local influence was part of an important debate inside the scientific community. In their article of 1935 entitled "Can quantum-mechanical description of physical reality be considered complete?", Einstein, Podolsky and Rosen [EPR35] argue that any theory making the same predictions as quantum theory and, at the same time, avoiding such *spooky action at a distance*, as they called these non-local influences, has to postulate the existence of "real properties" (now referred to as hidden-variables) which, when taken into account, allow for the complete local determination of the observations' outcomes. Since orthodox quantum theory does not include these, from the assumption of the impossibility of non-local causation one has to conclude its incompleteness. That same year, in an article with the same title, Niels Bohr argued otherwise. It took

almost 30 years for that discussion to substantially advance with Bell's 1964 celebrated result [Bel64], which states that the predictions of quantum mechanics for certain local measurements over spatially separated entangled systems cannot be accounted for by any local hidden-variable theory. Furthermore, Bell provided an experimental method to test whether Nature is, indeed, non-local or not [HBD+15, GVW+15, SMSC+15]: the violation of, what is now known as, a Bell inequality.

One of the assumptions in Bell's theorem is that the measurements in a Bell experiment are chosen independently of any state of affairs that may have any influence on the outcomes. This *measurement independence* assumption is easily satisfied by having the inputs be given by some random process. However, one may wonder:

> Is pseudorandomness in the choice of inputs to a Bell experiment enough to conclude the non-locality of the observed distribution of outcomes from a Bell inequality violation?

The motivation behind studying this question is, as usual, practical, since pseudorandomness is a cheap and readily available resource [MN98, Nie09], but also theoretical, since the only (theoretically) random processes we know are quantum measurements and one may not want to assume the validity of quantum mechanics in experiments designed to test the prediction of quantum mechanics.

1.1.2 Randomness vs pseudorandomness in the preparation of mixed states

Quantum mechanical systems can be in one of two types of states: pure or mixed. In the former, we have certainty about the state of the system and, therefore, the uncertainty about the measurement results is solely due to the indeterminism of the quantum measurement process. In the latter, which are probabilistic mixtures of pure states, an additional level of uncertainty is hence added. Continuing with the study of the implications of using pseudorandomness in quantum setups, we turn to the preparation of mixed states. Theoretically, one way to prepare a maximally mixed state of dimension n is by choosing uniformly at random states from a basis of $\mathbb{C}^n$. Hence, we ask ourselves:

> Does replacing the randomness source by a pseudorandom number generator (as done in e.g. the experiments of [AB09a, LKPR10]) in the preparation of mixed states have any observational consequences?

1.1.3 Randomness vs pseudorandomness in non-local hidden-variable theories

It is a consequence of Bell's theorem that any hidden-variable account of non-local correlations between the outputs of deterministic systems have to allow for local measurement outcomes to depend on distant measurement choices. That is, it has to violate what is usually known as *parameter independence* [Shi86]. On top of this "hidden-signaling mechanism", every deterministic account of quantum predictions given so far have also made use of some source of shared randomness [Boh52, TB03a]. In other words, the output of the systems in the n-th round of a Bell experiment is modelled as a function of the measurement choices (local and distant) plus some shared "hidden-variable" λ sampled from some distribution $p(\lambda)$. We consider the following question:

Does having the choice of hidden-variable at round n of a Bell experiment be, instead of random, a computable function of n have any physical consequence?

1.1.4 Randomness vs non-locality in communication complexity

Communication complexity studies the communication requirements of distributed computational tasks. Many tasks have been shown to be achievable through the communication of a number of qubits much smaller than the required number of bits, even when the distributed computing devices have access to some shared source of randomness [CvDNT99, Raz99b, BYJK04]. The similarity to the non-locality scenario naturally raises the following question:

Is non-locality the responsible when a function has quantum communication complexity smaller than classical?

1.2 Outline and contributions

In this thesis we use several tools from the theories of inductive inference, algorithmic randomness and communication complexity to address all the questions raised above, with the hope that our answers contribute to the progress in the understanding of the foundations of quantum mechanics.

In Chapter **??** we outline the basic concepts and definitions from theoretical computer science and quantum mechanics which will be needed in the following chapters.

In Chapter 2 we show that using pseudorandom number generators (PRNGs) in a Bell test opens up a loophole. In other words, we prove that if at least one of the players in a Bell test is using a PRNG to choose his inputs, then, under reasonable assumptions, local models for the observed correlations cannot be ruled out. The results of this chapter were published in [BdlTS$^+$16].

In Chapter 3 we give a protocol which, under reasonable assumptions, allows players Alice and Bob holding computable non-local boxes to signal faster than light. This implies that, since quantum correlations are non-signaling, any deterministic non-local hidden-variable account of quantum mechanics must have the assignment of hidden variables to rounds in a Bell test be uncomputable. This result will appear in [BdlTS$^+$17].

In Chapter 4 we show that there is a measurement strategy allowing a player Bob to distinguish any pseudorandom mixture of quantum states being prepared privately by another player Alice from the maximally mixed state, in finite time and with arbitrarily high success probability. Furthermore, we provide a proof-of-concept experiment of a special case of this result done by the group of Dr. Larotonda. These results appear in [BdlTS$^+$16] and [LGSdlT$^+$].

In Chapter 5 we show, for a large family of functions, how to construct Bell inequalities and quantum correlations violating them in a magnitude which is exponential in the difference between the quantum and classical communication complexities of the functions. Furthermore, we prove that the violations are resistant to the detection loophole and, with an increase in the number of outputs, can be also made resistant to uniform noise. These results appear in [LLN$^+$16]. We assume a basic knowledge of

computability theory and quantum mechanics, but we will outline all the key notions and concepts needed from these fields in this thesis.

1.3 Nomenclature and notation

We will denote by $\mathbb{N}$ the set of natural numbers. As usual we identify a set $A \subseteq \mathbb{N}$ with its characteristic function $\chi_A : \mathbb{N} \to \{0,1\}$ notated simply by A. That is, $A(x) = 1$ if $x \in A$ and $A(x) = 0$ if $x \notin A$. $\mathbb{R}$ is the set of reals and $\mathbb{C}$ the set of complex numbers. If $z \in \mathbb{C}$, then $|z|$ denotes its modulus. $\mathbb{Q}_2^{\geq 0} = \{n/2^m \mid n, m \in \mathbb{N}\}$ is the set of non-negative dyadic rationals. We represent $q \in \mathbb{Q}_2^{\geq 0}$ by a pair $\langle \sigma, \tau \rangle$ where σ and τ are binary strings in representing the integer and fractional part of q, respectively. We fix bijective functions taking a pair of natural numbers (a, b) to a natural number $\langle a, b \rangle$ and a sequence of natural numbers $(n_0, \ldots, n_k)$ to a natural number $[n_0, \ldots, n_k]$ (the only thing that matters here is that we can recover (a, b) from $\langle a, b \rangle$ and $(n_0, \ldots, n_k)$ from $[n_0, \ldots, n_k]$ in a computable way).

Let Σ be a finite set. Then, Σ^* is the set of all finite strings of symbols from Σ. ϵ denotes the empty string. If $b \in \{0,1\}$, then $\bar{b} := 1 - b$. We will be working with partial functions $f : \{0,1\}^* \to \{0,1\}^*$, that is, functions which can be undefined for some set $A \subseteq \{0,1\}^*$. We denote with $\operatorname{dom} f \subseteq \{0,1\}^*$ the set of inputs for which f is defined. If a partial function f is such that $\operatorname{dom} f = \{0,1\}^*$, then we say that f is total. For a string $\sigma \in \Sigma^*$, $|\sigma|$ denotes its length. For $i \in \{0, ..., |\sigma| - 1\}$, $\sigma(i)$ denotes the i-th symbol of σ. Sometimes we will consider natural numbers as binary strings. In this case, we use the string 0 to represent the natural number 0 and for any $n > 0$, we use the string that represents n in binary notation, starting with 1. Observe that when interpreting a number $n > 0$ as a string we have $|n| = 1 + \lfloor \log_2 n \rfloor$. Σ^ω denotes the set of all infinite sequences of symbols from Σ. For a set of strings $V \subseteq \Sigma^*$, $[V]$ denotes the set $\mathcal{V} \subseteq \Sigma^\omega$ of infinite sequences having a string in V as a prefix. The infinite sequence $A \in \{0,1\}^\omega$ can also be regarded as the enumeration $A(0)A(1)A(2)\ldots$ of the characteristic function of a set $A \subseteq \mathbb{N}$. Furthermore, A can also be seen as the real number in $[0,1]$ defined by $\sum_{n \geq 1} A(n) \cdot 2^{-n}$. We denote by $A \upharpoonright n$ the string of length n which consists of the first n symbols of A, that is, $A(0) \ldots A(n-1)$.

All vector spaces are assumed to be finite dimensional, unless otherwise noted. As it is customary in quantum mechanics, we will use Dirac's bra-ket notation. That is, $|\cdot\rangle$ will denote a column vector and $\langle \cdot|$ a row vector. If $|\psi\rangle$ and $|\phi\rangle$ are two vectors in $\mathbb{C}^n$ then $\langle \psi|\phi\rangle$ denote the usual inner product and $|\psi\rangle\langle\phi|$ the outer product, i.e. the linear operator over $\mathbb{C}^n$ defined as $(|\psi\rangle\langle\phi|)\,|\chi\rangle := \langle\phi|\chi\rangle|\psi\rangle$ for all $|\chi\rangle \in \mathbb{C}^n$. If A is a matrix, then A_{ij} denotes the element in row i and column j. We denote with $\operatorname{tr}(\cdot)$ the trace operation, i.e. given an operator A over $\mathbb{C}^n$, $\operatorname{tr}(A) = \sum_i A_{ii}$ for some matrix representation of A.

1.4 Quantum mechanics

1.4.1 States

In quantum mechanics, the *state space* of any physical system is modelled by a complex vector space with inner product (that is, a Hilbert space). The system is completely described by its *state vector*, which is a unit vector in the system's state space.

The simplest quantum mechanical system, and the system which we will be most concerned with, is the *qubit*. A qubit has a two-dimensional state space. Suppose $|0\rangle$ and $|1\rangle$ form an orthonormal basis for that state space, for example

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

Then, an arbitrary state vector in the state space can be written

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1.1}$$

with $\alpha, \beta \in \mathbb{C}$ and, by the unitary requirement, $|\alpha|^2 + |\beta|^2 = 1$.

Definition 1.4.1 (Qubit)**.** A *qubit* is a unit vector in $\mathbb{C}^2$.

Intuitively, the states $|0\rangle$ and $|1\rangle$ are analogous to the two values 0 and 1 which a classical bit may take. The way a qubit differs from a bit is that *superpositions* of these two states, i.e. linear combinations of $|0\rangle, |1\rangle$ with complex coefficients whose squared of moduli sum to one, can also exist. In a superposition, it *is not possible to say that the qubit is definitely in the state* $|0\rangle$*, or definitely in the state* $|1\rangle$.

Example 1. $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ form another orthonormal basis of $\mathbb{C}^2$.

Quantum mechanics also allows for the state of a system to be in a statistical mixture of m quantum states $|\psi_i\rangle, i = 1, \ldots, m$, each with probability p_i. We call $\{(p_i, |\psi_i\rangle)\}$ an *ensemble of states*, and say that a system with such an ensemble associated to it, is in a *mixed state*. In the formalism, these states are modelled through the use of a *density operator*, defined as

$$\rho := \sum_{i=1}^{m} p_i |\psi_i\rangle\langle\psi_i|. \tag{1.2}$$

The density operator is often known as the *density matrix*; we will use the two terms interchangeably. It is easy to see that ρ satisfies the following properties:

1. $\mathrm{tr}(\rho) = 1$
2. ρ is a positive operator.

Definition 1.4.2 (Mixed state)**.** A *mixed state* is a positive operator on a Hilbert space with unit trace.

Definition 1.4.3 (Maximally mixed state)**.** For every n, the *maximally mixed state* of dimension n is $\rho_n = \frac{1}{n}\mathbb{I}_n$, with $\mathbb{I}_n$ the $n \times n$ identity matrix.

For every basis $\{|\psi_i\rangle \mid 1 \leq i \leq n\}$ of $\mathbb{C}^n$, a system with an associate ensemble $\{(1/n, |\psi_i\rangle)\}$ is in a maximally mixed state. When a system is in a maximally mixed state, we have full uncertainty about its state. On the other hand, when we have certainty about the quantum state of a system, i.e. when $m = 1$, we say that the system is in a *pure state*. In other words, a pure state $|\psi\rangle$ can be seen as a special case of a mixed state with density matrix $|\psi\rangle\langle\psi|$.

Observation 1.4.4. *Two different ensembles can give rise to the same density matrix. For example, both the ensembles* $\{(1/2, |0\rangle), (1/2, |1\rangle)\}$ *and* $\{(1/2, |+\rangle), (1/2, |-\rangle)\}$ *give rise to the maximally mixed state* $\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

1.4.2 Observables and measurement

Measuring a quantum state is the quantum analogue of observing a classical state, but this measurement of a quantum state may modify it and this property is one of the most important features of quantum mechanics. It is impossible to directly observe the superposition of a quantum state; the only thing that we can do is to apply a measurement on the state in order to observe one of the classical states which belong to the superposition.

Measurable magnitudes in quantum mechanics are called *observables* and are modelled with Hermitian operators.

Definition 1.4.5 (Observable)**.** An *observable* is a Hermitian operator over a Hilbert space.

The possible outcomes obtainable when measuring an observable A are its eigenvalues, which, because of being a Hermitian operator, are real-valued. More formally, if A has spectral decomposition,

$$A = \sum_m m\Pi_m$$

with Π_m the projector onto the eigenspace associated with eigenvalue m, then the possible outcomes of measuring observable A correspond to the eigenvalues m. Upon measuring observable A to a system in a state ρ, the probability of getting result m is given by

$$p(m) = \mathrm{tr}(\Pi_m \rho),$$

and, after the measurement, the state of the measured system becomes

$$\rho' = \frac{\Pi_m \rho \Pi_m^\dagger}{\mathrm{tr}(\Pi_m \rho)}.$$

It is easy to see that the projectors $\{\Pi_m\}$ above satisfy

$$\sum_m \Pi_m = \mathbb{I}_n,$$

$$\Pi_i \Pi_j = \delta_{i=j}\Pi_i$$

where δ is the Kronecker delta function. A family of projectors $\{\Pi_m\}$ satisfying these identities describes a *projective measurement*.

Definition 1.4.6 (Pauli observables)**.** The Pauli observables, given here in their matrix form together with their spectral decomposition are

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = |+\rangle\langle+| - |-\rangle\langle-|,$$
$$\sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = (|0\rangle + i|1\rangle)(\langle 0| + i\langle 1|) - (|0\rangle - i|1\rangle)(\langle 0| - i\langle 1|),$$
$$\sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = |0\rangle\langle 0| - |1\rangle\langle 1|.$$

Any orthonormal basis $\mathbb{B} = \{|1\rangle, \ldots, |k\rangle\}$ can be associated to an observable $A_{\mathbb{B}}$ such that $\mathbb{B}$ is its eigenbasis,

$$A_{\mathbb{B}} = \sum_{i=1}^{k} a_i |i\rangle\langle i|.$$

Therefore, when we say that we *measure in the basis* $\mathbb{B}$ we mean measure observable $A_{\mathbb{B}}$.

Example 2. Measuring a system in state $|0\rangle$ in the eigenbasis of σ_x (i.e. $\{|+\rangle, |-\rangle\}$) leaves the system in state $|+\rangle$ or state $|-\rangle$ both with probability $1/2$. The same happens for every other combination of measuring on the eigenbasis of one Pauli operator a system whose state is an eigenstate of any of the other Pauli operators. The eigenbasis of σ_x, σ_y and σ_z are said to be *mutually unbiased*.

1.4.3 Multipartite systems and entanglement

When a physical system is made up of a number of smaller physical systems, we call it a *multipartite system*. If the composite system is n-partite, its Hilbert space $\mathcal{H}$ is given by the tensor product of the Hilbert spaces $\{\mathcal{H}_i\}_{1 \leq i \leq n}$ of the individual subsystems.

Definition 1.4.7 (Tensor product of vector spaces)**.** If V and V' are two vector spaces of dimension d and d' with respective basis $\{|v_1\rangle, \ldots, |v_d\rangle\}$ and $\{|v'_1\rangle, \ldots, |v'_{d'}\rangle\}$ then $V \otimes V'$ is the vector space of dimension $d \times d'$ spanned by $\{|v_i\rangle \otimes |v'_j\rangle \mid 1 \leq i \leq d, 1 \leq j \leq d'\}$.

We often use the abbreviated notations $|v\rangle|w\rangle$ or $|vw\rangle$ for the tensor product $|v\rangle \otimes |w\rangle$. Also, $|\psi\rangle^{\otimes n}$ will denote $|\psi\rangle$ tensored with itself n times.

Definition 1.4.8 (n-qubits register)**.** An n-qubits register is a unit vector in $(\mathbb{C}^2)^{\otimes n}$.

One can notice that if a vector in $V \otimes V'$ is a linear combination of vectors of the form $|v_i\rangle \otimes |v_j\rangle'$, it may not be possible to express it as the tensor product of a vector of V and a vector of V'. For example, $\sqrt{1/2}(|01\rangle - |10\rangle)$ is in $\mathbb{C}^2 \otimes \mathbb{C}^2$ and cannot be expressed as the tensor product of two vectors of $\mathbb{C}^2$. When a quantum state is in a vector space of the form $V \otimes V'$ and cannot be written as the tensor product of one

state of V and one state of V' we say that the subsystems of the state associated with V and V' are *entangled*.

When it comes to mixed states, the definition of entanglement changes a little bit. In this case, the entangled states are those which cannot be written as a convex combination of tensor products,

$$\rho = \sum_j p_j \otimes_{i=1}^n \rho_i^{(j)}.$$

1.5 Bell non-locality

Consider a bipartite system made up of two qubits in a joint state

$$|\psi^-\rangle := \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

and suppose that we give one qubit to Alice in lab A and another to Bob in lab B. According to quantum theory, if Bob were to measure observable σ_z to his qubit, he would obtain $+1$ or -1 with equal probability

$$\begin{aligned} p(+1) &= Tr\left((\mathcal{I} \otimes |0\rangle\langle 0|)|\psi\rangle\langle\psi|\right) = 1/2, \\ p(-1) &= Tr\left((\mathcal{I} \otimes |1\rangle\langle 1|)|\psi\rangle\langle\psi|\right) = 1/2. \end{aligned}$$

However, suppose that, first, Alice measures σ_z to her qubit. Quantum theory tells us that, if she obtains $+1$, the state of the composite system after the measurement is

$$\rho_{+1} = \frac{((\mathcal{I} \otimes |0\rangle\langle 0|)|\psi\rangle\langle\psi|(\mathcal{I} \otimes |0\rangle\langle 0|)}{p(+1)} = |01\rangle\langle 01|$$

and, if she obtains -1, the state becomes,

$$\rho_{-1} = \frac{((\mathcal{I} \otimes |1\rangle\langle 1|)|\psi\rangle\langle\psi|(\mathcal{I} \otimes |1\rangle\langle 1|)}{p(+1)} = |10\rangle\langle 10|.$$

Now, the result of measuring σ_z to Bob's qubit, which was completely uncertain in the first situation, becomes completely certain. That is, if Alice obtains $+1$, she *instantaneously* knows that Bob will obtain -1, because

$$p(-1) = Tr\left((\mathcal{I} \otimes |1\rangle\langle 1|)|01\rangle\langle 01|\right) = 1$$

and similarly if she obtains -1. Furthermore, this holds irrespectively of the distance separating labs A and B.

Einstein, Podolsky and Rosen [EPR35] proposed a *criterion of reality* by which: *if a property of a physical system can be determined without disturbing it, then such a property is an element of physical reality (EPR).* Next, they claimed that for a physical theory to be *complete*, it must assign values to all these EPRs. It follows that either,

- the act of measuring the qubit in lab A instantaneously disturb the state of affairs in lab B or

- it does not and so quantum theory is incomplete, because we can predict with certainty the value of measuring σ_z to Bob's qubit without disturbing it but for a system in the state $|\psi^-\rangle$ quantum theory does not assign a definite value to such physical property.

Therefore, if one does not want to abandon a *local* view of the Universe, as was the case of Einstein and co-authors, one must conclude that quantum theory was incomplete.

Almost 30 years later, Bell [Bel64] proved that quantum theory cannot be "completed" by a local theory. Furthermore, he provided an experimental method to test whether Nature is, as quantum theory predicts, non-local. In this section we develop the basics of the theory of Bell non-locality, necessary to understand the results that we present in the following chapters. Before going into that, however, let us note first that, although it serves to exemplify Einstein's discomfort with these non-local influences provided by entanglement (or, as he deemed them, this *spooky action at a distance*), there is nothing *unclassical* in the particular situation depicted above. Namely, if instead of sharing a pair of qubits in the state $|\psi^-\rangle$ and making the measurements just described, Alice and Bob received a box containing the result of flipping a fair coin, the situation would be identical: before Alice opens her box, there is uniform probability in the result of opening Bob's box; once Alice's box is opened, the result of opening Bob's box is determined. This is why EPR full argument involves two (non-commuting) observables. Next we will see an example of a truly quantum non-local situation.

1.5.1 Local, quantum and non-signaling distributions

The typical bipartite Bell scenario consists of two experimenters, Alice and Bob, each receiving a box with finite sets of inputs (the measurement choices), denoted $\mathcal{X}$ for Alice and $\mathcal{Y}$ for Bob, and finite sets of outputs (the measurement outcomes), $\mathcal{A}$ for Alice and $\mathcal{B}$ for Bob. The object of interest is

$$p(a, b|x, y),$$

the joint conditional probability distribution of observing outcomes $(a, b) \in \mathcal{A} \times \mathcal{B}$ when the inputs are $(x, y) \in \mathcal{X} \times \mathcal{Y}$. More formally, we consider bipartite distribution families of the form $\mathbf{p} = (p(\cdot, \cdot|x, y))_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ with inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ determining a probability distribution $p(\cdot, \cdot|x, y)$ over the outcomes $(a, b) \in \mathcal{A} \times \mathcal{B}$, with the usual positivity and normalization constraints. For simplicity, we call simply "distributions" such probability distribution families. We will use the expression "Alice's marginal" to refer to her marginal output distribution, that is $\sum_b p(\cdot, b|x, y)$ (and similarly for Bob).

The first kind of distributions we will consider are the *local-deterministic* ones.

Definition 1.5.1 (local-deterministic distributions)**.** A distribution $\mathbf{p}$ is *local-deterministic* iff

$$p(a, b|x, y) = \delta_{a=\lambda_A(x)}\delta_{b=\lambda_B(y)}$$

where λ_A (resp. λ_B) is a function from $\mathcal{X}$ to $\mathcal{A}$ (resp. from $\mathcal{Y}$ to $\mathcal{B}$). We denote by $\mathcal{L}_{det}$ this set of distributions.

Next, by taking convex combinations we have a (geometrical) definition of the set $\mathcal{L}$ of local distributions.

Definition 1.5.2 (local distributions)**.** $\mathcal{L}$ is the convex-hull of $\mathcal{L}_{det}$. That is, $\mathbf{p} \in \mathcal{L}$ iff

$$p(a,b|x,y) = \sum_{\ell \in \mathcal{L}_{det}} p(\ell)\ell(a,b|x,y).$$

Equivalently, the local distributions are those admitting a *local λ-independent hidden-variable model*.

Definition 1.5.3 (hidden-variable model)**.** A *hidden-variable model* for a distribution $\mathbf{p}$ is a tuple $\langle \Lambda, q, \mathbf{p}_\lambda \rangle$ such that

1. Λ is a finite set (the hidden-variables),
2. $q : \Lambda \to \mathbb{R}$ is a probability distribution (the shared randomness).
3. $\mathbf{p}_\lambda = (p_(\cdot,\cdot|x,y,\lambda))_{(x,y)\in\mathcal{X}\times\mathcal{Y}\times\Lambda}$ is a family of probability distributions over $\mathcal{A}\times\mathcal{B}$ (the boxes) of which we assume that all of their marginals are well defined.
4. $p(a,b|x,y) = \sum_\lambda q(\lambda|x,y)p(a,b|x,y,\lambda)$

Definition 1.5.4 (λ-independent)**.** We say that a hidden-variable model $\langle \Lambda, q, \mathbf{p}_\lambda \rangle$ is *λ-independent* iff $q(\lambda|x,y) = q(\lambda)$ for all $(x,y,\lambda) \in \mathcal{X}\times\mathcal{Y}\times\Lambda$.

Definition 1.5.5 (local hidden-variable model)**.** We say that a hidden-variable model $\langle \Lambda, q, \mathbf{p}_\lambda \rangle$ is *local* iff

$$p(a,b|x,y,\lambda) = p(a|x,\lambda)p(b|y,\lambda). \tag{1.3}$$

Definition 1.5.6 (deterministic hidden-variable model)**.** A *deterministic hidden-variable model* $\langle \Lambda, q, A, B \rangle$ is a hidden-variable model $\langle \Lambda, q, \mathbf{p}_\lambda \rangle$ such that

$$p(a|x,y,\lambda) = \delta_{a=A(x,y,\lambda)} \text{ and } p(b|x,y,\lambda) = \delta_{b=B(x,y,\lambda)}$$

where $A : \mathcal{X}\times\mathcal{Y}\times\Lambda \to \mathcal{A}$ and $B : \mathcal{X}\times\mathcal{Y}\times\Lambda \to \mathcal{B}$.

Proposition 1.5.7. *The local distributions $\mathcal{L}$ are those admitting a local deterministic λ-independent hidden-variable model.*

Fine [Fin82] showed that restricting to deterministic models is without loss of generality. Intuitevely, this is because if they were indeterministic we could "factor out" that indeterminism into the shared randomness.

Theorem 1.5.8 ([Fin82])**.** *A distribution $\mathbf{p}$ has a local hidden-variable model iff it has a local deterministic hidden-variable model.*

Operationally, local distributions are those that can arise from pairs of (non– communicating) deterministic boxes with access to some shared randomness. Equivalently, this means that the state of the whole experiment at the moment of performing the measurement is described by some past common cause (usually referred to as *hidden-variable*) λ such that, to describe the statistics, one averages over all the possible states of the experiment according to some distribution q. The boxes produce their outputs solely according to the information locally available to them: the (local) input and the hidden variable λ.

The *quantum distributions* are those that arise from Alice and Bob making local measurements over a biparite quantum state.

Definition 1.5.9 (quantum distributions)**.** A distribution $\mathbf{p}$ is *quantum* if there exist a density operator ρ over a joint Hilbert space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ of arbitrary dimension, a family $\{A^{(x)}\}_{x\in\mathcal{X}}$ of $|\mathcal{A}|$-outcomes observables $A^{(x)} = \sum_{a\in\mathcal{A}} a\Pi_a^{(x)}$ over $\mathcal{H}_\mathcal{A}$ and a family $\{B^{(y)}\}_{y\in\mathcal{Y}}$ of $|\mathcal{B}|$-outcomes observables $B^{(y)} = \sum_{b\in\mathcal{B}} b\Pi_b^{(y)}$ over $\mathcal{H}_\mathcal{B}$ such that $p(a,b|x,y) = \mathrm{tr}((\Pi_a^{(x)} \otimes \Pi_b^{(y)})\rho)$. We denote by $\mathcal{Q}$ this set of distributions.

Finally, a distribution is *non-signaling* if for each player, its marginal output distributions do not depend on the other player's input.

Definition 1.5.10 (non-signaling distributions)**.** A distribution $\mathbf{p}$ is non-signaling iff

$$\sum_b p(a,b|x,y) = \sum_b p(a,b|x,y') \text{ and}$$
$$\sum_a p(a,b|x,y) = \sum_a p(a,b|x',y)$$

We denote by $\mathcal{NS}$ this set of distributions.

These constraints have a clear physical interpretation: they imply that the local marginal probabilities of Alice $p(a|x) := \sum_b p(a,b|x,y)$ are independent of Bob's measurement setting y, and thus Bob cannot signal to Alice by his choice of input (and the other way around).

Example 3 (PR-box)**.** Popescu and Rohrlich [PR94] showed that the distribution

$$p(a,b|x,y) = \begin{cases} 1/2 & \text{if } a \oplus b = xy \\ 0 & \text{otherwise} \end{cases}$$

with inputs in $\mathcal{X} \times \mathcal{Y} = \{0,1\}^2$ and outputs in $\mathcal{A} \times \mathcal{B} = \{0,1\}^2$ and now known as a *PR-box*, is non-signaling but also non-quantum.

1.5.2 Bell inequalities

Once we have a definition of what a local distribution is, it is not hard to see that some probability distributions arising in quantum theory are not local. The way we do this is by coming up with a *Bell inequality*, that is, a linear constraint

$$B = \sum_{a,b,x,y} B_{a,b,x,y} p(a,b|x,y) \leq B_L$$

over distributions $\mathbf{p}$ which is satisfied by every local distribution, but which can be violated by quantum distributions.

Example 4. For ease of presentation, let us consider the simplest scenario: two inputs per party $x, y \in \{0,1\}$ with two possible outputs $a, b \in \{-1,+1\}$ each. In this setting, there is a unique (up to relabelling of the inputs and outputs) Bell inequality that

is tight on the set of local probabilities: the Clauser-Horne-Shimony-Holt (CHSH) inequality [CHSH69, Fin82]

$$B := \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2, \tag{1.4}$$

where $\langle A_x B_y \rangle = \sum_{a,b} abp(a,b|x,y)$. It is easy to see that if Alice and Bob are given pairs of qubits in the singlet state $|\Psi^-\rangle = (1/\sqrt{2})(|01\rangle - |10\rangle)$ and we let $x, y \in \{0,1\}$ label observables,

$$A^{(0)} = \sigma_x \qquad B^{(0)} = \frac{-\sigma_z - \sigma_x}{\sqrt{2}}$$
$$A^{(1)} = \sigma_z \qquad B^{(1)} = \frac{\sigma_z - \sigma_x}{\sqrt{2}}$$

the value of B for the resulting quantum distribution is

$$B = 2\sqrt{2} > 2. \tag{1.5}$$

which we know from is the maximum achievable by any quantum distribution.

Tsirelson [Cir80] showed that $2\sqrt{2}$ is the maximum violation of the CHSH inequality achievable by quantum distributions. This, together with the fact that the PR-box achieves a value of 4 (which is, also, the algebraic maximum) gives us the well know inclusion between the classes of distributions presented before:

$$\mathcal{L} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}.$$

1.5.3 Loopholes

Violations of Bell inequalities have been observed experimentally in a variety of physical systems, giving strong evidence that nature is indeed, as predicted by quantum theory, non-local [ADR82, TBZG98, WJS+98, RKM+01]. However, it was not until 2015 that the first experiments considered *loophole-free* were performed [HBD+15, GVW+15, SMSC+15].

Definition 1.5.11 (loophole)**.** A loophole, in the context of Bell experiments, is an experimental situation allowing classical devices to generate non-local correlations.

The experiments of [HBD+15, GVW+15, SMSC+15] were the first to simultaneously close the famous *locality* and *detection* loopholes.

Locality loophole.

The definition of locality (Definition 1.5.2) is motivated by the absence of communication between the measurement sites of a Bell experiment. This seems well justified if the sites are sufficiently separated so that the time elapsed between the decision of which measurement to make is made and the measurement output is recorded is shorter than the time taken by a signal travelling at the speed of light, to travel from one site to another. If this condition is not satisfied, one could in principle conceive a purely "local" mechanism (i.e., involving slower-than-light speed signals) underlying the observed correlations.

Detection loophole

In a large class of Bell experiments, in particular those carried out with photons, measurements do not always yield conclusive outcomes. This is due either to losses between the source of particles and the detectors or to the fact that the detectors themselves have non-unit efficiency. In this scenario, in addition to the outcomes in $\mathcal{A}\times\mathcal{B}$, we have two additional "no-click" outcomes per side, denoted $\perp$. If in a Bell experiment, we discard the "no-click" rounds and the remaining rounds do not comprise a *fair-sampling* of all rounds because the detectors were somehow coordinating their behaviour by deliberately choosing when not to click, it could happen that though the conditional probability (conditioned on neither outcome being $\perp$) may look quantum, the unconditional probability may very well be classical (local). This is called the detection loophole.

Example 5 (Exploiting the detection loophole.)**.** To illustrate the idea, let us see how to locally violate the CHSH inequality by exploiting the detection loophole. The local model is as follows. The hidden-variable λ corresponds to two uniform random bits x_{guess} and a. Given measurement setting y, Bob's detector outputs $b = a \oplus x_{guess}y$. Alice's detectors output a whenever her measurement setting is $x = x_{guess}$ and output $\perp$ when $x \neq x_{guess}$. Focusing on the conclusive outcomes (i.e. ± 1), the value of the CHSH Bell functional B (1.4) for the conditional distribution is 4. The probability for Alice to obtain a conclusive outcome is $1/2$, which is the probability that $x = x_{guess}$, while Bob always obtains a conclusive outcome. With additional shared randomness, it is possible to symmetrize the above model, such that Alice and Bob's detection probability is $2/3$ [MP03a]. Therefore, if the detection efficiency in a CHSH Bell experiment is below $2/3$, no genuine Bell inequality violation can be obtained, since the above local strategy could have been used by the measurement apparatuses.

There are two ways to close the detection loophole: 1) work with detectors with high enough efficiency (see e.g. [Ebe93, MP03b, VPB10] for the minimum efficiency required in different scenarios) or 2) consider $\perp$ as a valid outcome and study the violation of inefficiency-resistant Bell inequalities [MP03b]. In Chapter 5 we will encounter these inequalities which appear as solutions to lower bound techniques in communication complexity [LLR12].

All these definitions and results about the theory of Bell non-locality, although sufficient for the purposes of this thesis, are only a small part of a vast theory which, for example, considers general multipartite scenarios with not necessarily symmetric number of inputs and outputs per site, and other generalizations of the like. For a thorough and complete reference, see [BCP$^+$14].

1.6 Computability theory

Part of the philosophy underlying computability theory is the celebrated Church-Turing thesis, which states that *the algorithmic (i.e., intuitively computable) functions are exactly those that can be computed by the formal model of Turing machines [Tur37].* Informally, a Turing machine (TM) $\mathbf{M}$ is just a computer program used to perform a specific task: it takes a binary string s as input and either gets undefined (notated

$\mathbf{M}(s)\uparrow$) or it halts (notated $\mathbf{M}(s)\downarrow$) and produces a certain binary string w as output. In this last case we say that $\mathbf{M}(s) = w$.

Definition 1.6.1. A partial function $f : \{0,1\}^* \to \{0,1\}^*$ is *partially computable* iff there exists a Turing machine $\mathbf{M}$ such that, for every $s \in \operatorname{dom} f$, $\mathbf{M}(s)\downarrow$ and $\mathbf{M}(s) = f(s)$, and, for every $s \notin \operatorname{dom} f$, $\mathbf{M}(s)\uparrow$. We say that $\mathbf{M}$ *partially computes* f.

Recall that if $\operatorname{dom} f = \{0,1\}^*$, we say that f is total. We will denote with $\mathcal{P}$ the class of all partial computable functions and with $\mathcal{R} \subseteq \mathcal{P}$ the class of total computable functions.

Looking only at $\{0,1\}^*$ may seem rather restrictive. For example, later we will be concerned with functions that take natural numbers or subsets of the rationals as their domains and/or ranges. However, from the point of view of computability theory (that is, where resources such as time and memory do not matter), our definitions naturally extend to such functions via coding; that is, the domains and ranges of such functions can be coded as subsets of $\{0,1\}^*$. Henceforth, unless otherwise indicated, when we discuss computability issues relating to a class of objects, we will always regard these objects as (implicitly) computably coded in some way.

Any Turing machine $\mathbf{M}$ can be approximated step by step. By $\mathbf{M}_t(s)\downarrow = w$ we denote that the machine $\mathbf{M}$ on input s halts within t computational steps and outputs w; by $\mathbf{M}_t(s)\uparrow$ we denote that $\mathbf{M}$ has not reached a halting state by stage t. At each stage t we can algorithmically determine if $\mathbf{M}$ has reached a halting state or not: if $\mathbf{M}_t(s)\downarrow$ then $\mathbf{M}_{t'}(s)\downarrow = \mathbf{M}(s)$ for all $t' \geq t$.

Definition 1.6.2. Let $f : \{0,1\}^* \to \{0,1\}^*$ and $T : \mathbb{N} \to \mathbb{N}$ be some total functions. We say that f is *computable in $O(T(n))$-time* iff there exists a Turing machine $\mathbf{M}$ computing f and a constant c such that for almost all $s \in \{0,1\}^*$, $\min_t[M_t(s)\downarrow] \leq c \cdot T(|s|)$.

Nowadays the fact that programs can be treated as strings and that all of them can be listed is quite standard for a computer scientist. Moreover, there are special programs that take strings representing programs as input, and simulate them. This is a consequence of the Enumeration Theorem, which says that we can algorithmically enumerate all the Turing machines

$$\mathbf{T}_0, \mathbf{T}_1, \mathbf{T}_2, \ldots \tag{1.6}$$

and that there is a universal machine $\mathbf{V}$ such that

$$\mathbf{V}(0^i 1 s) = \mathbf{T}_i(s), \tag{1.7}$$

so $\mathbf{V}$ can simulate every other machine. Each Turing machine $\mathbf{M}$ corresponds to a number in the list (1.6) and that number is called the Gödel number of $\mathbf{M}$. We will use the same symbol $\mathbf{T}_e$ for denoting the e-th Turing machine (regarded as a computing agent) and the (mathematical) function it computes, interchangeably.

The Recursion Theorem due to Kleene [Kle38] says that for every computable function f we can compute an n such that $\mathbf{T}_n = \mathbf{T}_{f(n)}$. This n is usually called the *fixed point* of f. This result (together with the *s-m-n* Theorem, see [Soa99] for more details) allows us to define computable functions which *know* in advance its own Gödel number.

For example, when defining a machine $\mathbf{M}$ we may assume that we already know the number e such that $\mathbf{M} = \mathbf{T}_e$.

Computability theory is usually concerned with sets.

Definition 1.6.3. A set is $B \subseteq \{0,1\}^*$ is *computable* iff its characteristic function χ_B is a computable function.

Definition 1.6.4. A set $B \subseteq \{0,1\}^*$ is called *computably enumerable* (c.e.) iff it is either empty or the range of some computable function $g : \mathbb{N} \to \{0,1\}^*$.

The notion of computable enumeration gives us an equivalent characterization of computable sets.

Proposition 1.6.5. *A set $B \subseteq \{0,1\}^*$ is* computable *iff both itself and its complement are c.e.*

The famous Undecidability of the Halting Problem, due to Turing [Tur37], says that

Theorem 1.6.6. *The set $\{e \in \mathbb{N} \mid \mathbf{T}_e(e)\downarrow\}$ is c.e. but not computable.*

The notion of a computable enumeration extends to classes of computable functions as follows:

Definition 1.6.7. A class $\mathcal{C}$ of partially computable functions is *computably enumerable* iff there exists a total computable function $g : \mathbb{N} \to \mathbb{N}$ such that

1. for every n, $\mathbf{T}_{g(n)} \in \mathcal{C}$ and
2. for every $f \in \mathcal{C}$, there exists $n \in \mathbb{N}$ such that $\mathbf{T}_{g(n)} = f$.

As a consequence of the Enumeration Theorem, we have the following result which will be important for what comes next,

Theorem 1.6.8. *For every computable $T : \mathbb{N} \to \mathbb{N}$, the class of functions computable in $O(T(n))$-time is computably enumerable.*

1.7 Inductive inference

The process of inductive inference can be described as a particular step from chaos (a sequence of accidents) to order (a pattern), or from effects (a sequence of events) to causes (a possible explanation of what produces them). It consists of generating hypotheses for describing an unknown object from finitely many data points about the unknown object. For example, when exploring a physical phenomenon by performing experiments, a physicist obtains a finite sequence of pairs

$$(x_0, f(x_0)), (x_1, f(x_1)), \ldots, (x_n, f(x_n)).$$

From these examples the physicist tries to infer the law f describing the connection between x and $f(x)$. Usually f is a mathematical expression, a formula, or, in a very general scenario, an *algorithm* computing the function f. Using more and more

examples, the hypothesis on hand may be confirmed or falsified. If it is falsified, usually a new hypothesis is generated.

Gold[Gol67] considers inductive inference to be an infinite process. The objects to be inferred are computable functions. In every step $n = 0, 1, 2, \ldots$ of the inference process the inference algorithm has access to successively growing initial segments $(x_0, f(x_0)), (x_1, f(x_1)), \ldots, (x_n, f(x_n))$ of the graph of the target function. Using these initial segments, the inference algorithm computes hypotheses h_n which are interpreted as numbers of programs in a given computable numbering of (all) partial computable functions (e.g. (1.6)).

Trivially, for every computable function there is an inference algorithm (namely, the one that always outputs the index of a program computing the function) and so, every computable function is *individually* inferable. We can thus turn our attention to *classes* of computable functions. The problem here is that, for each class $\mathcal{C}$ of computable functions, one could be able to infer members of $\mathcal{C}$ individually, without having a master inference method that would work uniformly for all members of $\mathcal{C}$.

Many possible formalizations of notions of inference for classes of total computable functions have been considered in the literature. We confine ourselves here to a few of them, those which we will be using in the next chapters, and refer to [ZZ08] for surveys of many others, as well as for bibliographical references on them. We state the results for functions $\mathbb{N} \to \mathbb{N}$ but, as usual, they easily extend to functions over other countable sets.

Before proceeding to the definitions, let us state some notation. The set of all permutations of $\mathbb{N}$ is denoted by $\mathrm{Perm}(\mathbb{N})$. Any element $X \in \mathrm{Perm}(\mathbb{N})$ can be represented by a unique sequence $(x_n)_{n\in\mathbb{N}}$ that contains each natural number precisely once. Let $X \in \mathrm{Perm}(\mathbb{N})$, $f : \mathbb{N} \to \mathbb{N}$ and $n \in \mathbb{N}$. Then we write $f^{X,n}$ instead of $[\langle x_0, f(x_0)\rangle, \ldots, \langle x_n, f(x_n)\rangle]$. If $X = 0, 1, 2, \ldots$, $f^n := f^{X,n}$.

1.7.1 Learnability in the limit

First, let us state the formal definition of Gold's original model of *learnability in the limit*.

Definition 1.7.1 ([Gol67])**.** A class $\mathcal{C}$ of total computable functions is *learnable in the limit* iff there exists a total computable function $g : \mathbb{N} \to \mathbb{N}$ (called a *learner* for $\mathcal{C}$) such that, for all $X \in \mathrm{Perm}(\mathbb{N})$ and every $f \in \mathcal{C}$ there exist $n_f \in \mathbb{N}$,

$$\mathbf{T}_{g(f^{X,n_f})} \text{ computes } f \text{ and}$$
$$g(f^{X,n}) = g(f^{X,n_f}) \text{ for all } n \geq n_f.$$

We denote with $\mathcal{LIM}$ the set of all this classes.

Notice that a finite number of wrong hypothesis (indexes of TMs) for each element of the class is allowed (i.e. g can take guesses and learn from its mistakes).

For our applications in Chapters 2 and 3 we will need that the TMs hypothesized by the learner always halt. This is formalized as follows.

Definition 1.7.2 ([Wie78])**.** A class $\mathcal{C}$ of total computable functions is $\mathcal{R}$-*totally-learnable* iff there exists a learner $g : \mathbb{N} \to \mathbb{N}$ for $\mathcal{C}$ such that $\mathbf{T}_{g(n)}$ is total for all n. We denote with $\mathcal{R}$-$\mathcal{TOTAL}$ the set of all this classes.

Observation 1.7.3. $\mathcal{R}$-$\mathcal{TOTAL} \subseteq \mathcal{LIM}$.

It follows from a simple diagonalization argument that,

Theorem 1.7.4. *The class of all total computable functions $\mathcal{R}$ is not in $\mathcal{R}$-$\mathcal{TOTAL}$.*

Proof. By way of contradiction, suppose that $\mathcal{R} \subseteq \mathcal{R}$-$\mathcal{TOTAL}$ via the learner $g : \mathbb{N} \to \mathbb{N}$ and let $f : \mathbb{N} \to \mathbb{N}$ be defined as

$$f(0) := 0 \quad \text{and} \quad f(n+1) := \mathbf{T}_{g(f^n)}(n+1) + 1.$$

Then f is total computable function such that $f(n+1) \neq \mathbf{T}_{g(f^n)}(n+1)$ for all n and so not learnable by g, a contradiction. □

On the other hand, large classes of total computable functions are $\mathcal{R}$-totally-learnable. In particular, we have the following characterization:

Theorem 1.7.5 ([AB91])**.** *A class of total computable functions is in $\mathcal{R}$-$\mathcal{TOTAL}$ if and only if it is a subclass of a computably enumerable class of total computable functions.*

Proof. We prove the *if* direction because it is instructive for the results of the following chapters. Let $h : \mathbb{N} \to \mathbb{N}$ be a computable enumeration of $\mathcal{C}$. Then $g : \mathbb{N} \to \mathbb{N}$, defined as follows,

$$g([\langle a_0, b_0\rangle, \ldots, \langle a_n, b_n\rangle]) := h(\min_{m \leq n}[\forall i \leq n \ \mathbf{T}_{h(m)}(a_i) = b_i]) \tag{1.8}$$

is a learner for $\mathcal{C}$ and, for every n, $\mathbf{T}_{g(n)}$ computes a total function. Hence, $\mathcal{C} \in \mathcal{R}$-$\mathcal{TOTAL}$. This is an example of a technique known as *learning by enumeration*. By assumption, for every $f \in \mathcal{C}$, there is (at least) one $n \in \mathbb{N}$ such that $f = \mathbf{T}_{h(n)}$. The predictor's guess on input $f^{X,n} = [\langle x_0, f(x_0)\rangle, \ldots, \langle x_n, f(x_n)\rangle]$ is the first program in the enumeration whose outputs on inputs $x_0, \ldots, x_n$ coincide with $f(x_0), \ldots, f(x_n)$ respectively. After finitely many mistakes, it will reach the first program in the enumeration computing f and, hence, its guesses will start being correct from then onwards. See Figure 1.1 for an schematic description. □

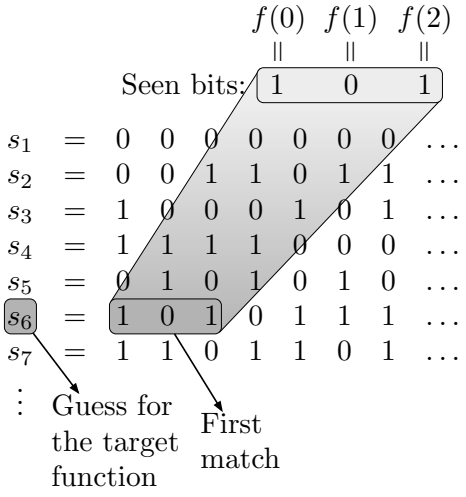

Fig. 1.1: Let $\{s_i\}_{i\in\mathbb{N}}$ be a computable enumeration of a class of (in this case $0,1$-valued) total computable functions. Learning by enumeration works as folllows: after seeing $f(0)=1$, $f(1)=0$ and $f(2)=1$, the guess for (a program comoputing) f will be the first TM in the enumeration whose outputs match those values (in the example, s_6.)

Definition 1.7.6 (enumeration-learner)**.** Let $\mathcal{C}$ be a computably enumerable class of total computable functions. For every function $h:\mathbb{N}\to\mathbb{N}$ enumerating $\mathcal{C}$ we will call a function $g:\mathbb{N}\to\mathbb{N}$ defined as in equation (1.8) an *enumeration-learner* for $\mathcal{C}$.

From Theorems 1.6.8 and 1.7.5 we have,

Corollary 1.7.7. *For every computable $T:\mathbb{N}\to\mathbb{N}$, the class of functions computable in $O(T(n))$-time is in $\mathcal{R}$-$\mathcal{TOTAL}$.*

This implies that the well-known classes **P**, **BQP**, **NP**, **PSPACE** from complexity theory, where the complexity time bound T is a simple exponential function and the much broader class **PR** of the primitive recursive functions where the time bound is Ackermannian [Odi92, §VIII.8] are all *learnable*.

For the application in Chapter 3 it will be sufficient for us that the learner outputs TMs coinciding with the target function in all but finitely many inputs. This has been formalized in the literature as follows:

Definition 1.7.8. A class $\mathcal{C}$ of total computable functions is *learnable in the limit with finite anomalies* iff there exists a total computable function $g:\mathbb{N}\to\mathbb{N}$ such that, for for all $X\in\mathrm{Perm}(\mathbb{N})$ and every $f\in\mathcal{C}$ there exist $n_f\in\mathbb{N}$,

$$g(f^{X,n})=g(f^{X,n_f}) \text{ for all } n\geq n_f \text{ and}$$
$$\exists m_0\forall m\geq m_0\ \mathbf{T}_{g(f^{X,n_f})}(m)=f(m).$$

We denote with $\mathcal{LIM}^*$ the set of all this classes.

1.7.2 Identification by next value

The other notion we consider formalizes the idea of a uniform method of prediction or extrapolation.

Definition 1.7.9. A class $\mathcal{C}$ of total computable functions is *identifiable by next value* iff there exists a total computable function $g : \mathbb{N} \to \mathbb{N}$ (called a *predictor* for $\mathcal{C}$) such that, for every $f \in \mathcal{C}$ there exists $n_f \in \mathbb{N}$,

$$f(n+1) = g(f^n) \text{ for all } n \geq n_f.$$

We denote with $\mathcal{NV}$ the set of all this classes.

Again, notice that a finite number of wrong predictions for each element of the class is allowed.

In Chapter 2 we will use this model because it simplifies the explanation but, it can be shown that,

Theorem 1.7.10 (see e.g. Theorem 2 of [ZZ08])**.** $\mathcal{NV} = \mathcal{R}\text{-}\mathcal{TOTAL}$

And, together with Corollary 1.7.7 it follows that

Corollary 1.7.11. *For every computable $T : \mathbb{N} \to \mathbb{N}$, the class of functions computable in $O(T(n))$-time is in $\mathcal{NV}$.*

Ideally, one would like to be able to effectively tell when the predictions begin to be correct. One way to formalize this is as follows,

Definition 1.7.12. A class $\mathcal{C}$ of total computable functions is *finitely identifiable by next value* iff there exists a computable function $g : \mathbb{N} \to \mathbb{N}$ (called a *finite-predictor* for $\mathcal{C}$) such that for all $f \in \mathcal{C}$, there exists an n such that,

$$g(f^m) = \langle \cdot, 0 \rangle \text{ for all } m < n \text{ and}$$
$$g(f^m) = \langle f(m+1), 1 \rangle \text{ for all } m \geq n.$$

We denote with $\mathcal{NV}_{fin}$ the set of all this classes.

Of course, $\mathcal{NV}_{fin} \subseteq \mathcal{NV}$. But, unfortunately, very simple classes are already outside $\mathcal{NV}_{fin}$.

Example 6. Consider the class $\mathcal{C} = \{f_n \in \mathcal{R} \mid f_n(n) = n \wedge f_n(x) = 0 \text{ if } x \neq n\}$. Straightforwardly, $\mathcal{C} \in \mathcal{NV}$ via, e.g., the predictor constantly 0. Suppose $\mathcal{C} \in \mathcal{NV}_{fin}$ and let g be a finite-predictor for $\mathcal{C}$. Since the always 0 function is in $\mathcal{C}$, there exists n_0 such that $g(0^n) = \langle 0, 1 \rangle$ for all $n \geq n_0$. But then g doesn't predict $f_{n_0+1} \in \mathcal{C}$, because $g(f^{n_0}_{n_0+1}) = g(f^{n_0}_0) = \langle 0, 1 \rangle \neq \langle f_{n_0+1}(n_0+1), 1 \rangle$. A contradiction. Note that every $f \in \mathcal{C}$ is computable in $O(\log n)$-time; thus, it takes very little time complexity to go outside of $\mathcal{NV}_{fin}$.

1.8 Computable randomness

In the preceding section, we said that we were going to consider the predictability of classes of computable sequences, instead of individual computable sequences, because the latter are trivially predictable. What about uncomputable sequences? Consider the sequence (equivalently, the set),

$$K := \{e : \mathbf{T}_e(e) \downarrow\}.$$

We saw in Theorem 1.6.6 that K is not computable and, so, of course, after seeing $K \upharpoonright n$, for any n, we will not be able to predict the forthcoming bits (in the sense of the preceding section). However, because it is c.e., we can certainly program a Turing machine $\mathbf{M}$ to, given $K \upharpoonright n$, tell us a position $m \geq n$ where $K(m) = 1$, that is

$$\text{if } \mathbf{M}(K \upharpoonright n) = m \quad \text{then} \quad m \geq n \text{ and } K(m) = 1. \tag{1.9}$$

Therefore, in what can be seen as a weaker sense of *predictability*, K, although uncomputable, is predictable.

The notion of *algorithmic unpredictability* is formalized through the use of *martingales*. Consider the following betting game. A gambler in a casino is presented with larger and larger bits of a binary sequence

$$X = X(0)X(1)X(2)\dots$$

His initial capital is $M(\sigma) \geq 0$. After seeing $x = X(0)\dots X(k-1)$, his capital is $M(x) \geq 0$. He places a bet for the value of the $(k+1)$-th bit as follows: he bets $m(x0) \geq 0$ to $X(k)$ being 0 and $m(x1) \geq 0$ to it being one, with $m(x0)+m(x1) \leq M(x)$. Then, $X(k) = b$ is revealed and he wins $2 \cdot m(xb)$. His new capital thus becomes $M(X(0)\dots X(k)) = M(x) - (m(x0) + m(x1)) + 2m(xb)$. The rules of the game are fair in the sense that the expected capital after each bit is equal to the current capital, that is

$$\frac{M(x0) + M(x1)}{2} = M(x). \tag{1.10}$$

Formally, and for any finite set of symbols Σ,

Definition 1.8.1 (Martingale)**.** A function $M : \Sigma^* \to \mathbb{Q}_2^{\geq 0}$ is called a *martingale* iff

$$\frac{1}{|\Sigma|} \sum_{b \in \Sigma} M(\sigma b) = M(\sigma). \tag{1.11}$$

Recall that $\mathbb{Q}_2^{\geq 0} = \{n/2^m \mid n, m \in \mathbb{N}\}$ denotes the set of non-negative dyadic rationals.

Martingales formalize the notion of betting strategy. At each step, the gambler must bet $\frac{M(\sigma b)}{|\Sigma| M(\sigma)}$ of his current capital to the next bit being b. The betting strategy is successful when the gambler's capital increases unboundedly when playing according to the strategy on the successive symbols of X.

Definition 1.8.2 (Success of a martingale)**.** We say that a martingale M succeeds on a sequence $X \in \Sigma^\omega$ iff $\limsup_n M(X \upharpoonright n) = \infty$.

As the reader might be expecting, we will be considering betting strategies of an *effective* kind.

Definition 1.8.3 (Computable randomness)**.** A sequence $X \in \Sigma^\omega$ is *computably random* if no computable martingale succeeds on it. We denote with CR the set of computably random sequences.

Of course, no computable sequence is computably random. Furthermore, to succeed on a sequence X it suffices being able to effectively *pinpoint* a non computably random subsequence.

Proposition 1.8.4. *Let $Z \in \Sigma^\omega$ and $h : \mathbb{N} \to \mathbb{N}$ be strictly increasing and computable. If Z is computably random, then $X = Z(h(0))Z(h(1))Z(h(2))\dots$ is computably random.*

Proof. By contrapositive, let M be a computable martingale that succeeds on X. It is easy to see that the martingale $F : \Sigma^* \to \mathbb{Q}_2^{\geq 0}$ defined as

$$F(\epsilon) = M(\epsilon)$$
$$F(\sigma b) = \begin{cases} M(\sigma[h(0)]\dots\sigma[h(n-1)]b) & \text{if } h(n) = |\sigma| \\ F(\sigma) & \text{otherwise} \end{cases}$$

is computable and succeeds on Z. □

For the results of Chapter 3 we will need computable sets which are *sufficiently random*. This is formalized with the notion of *resource-bounded randomness* in which the computational power of the *gambler* is restricted.

Definition 1.8.5 ($T(n)$-randomness)**.** Let $T(n) : \mathbb{N} \to \mathbb{N}$ be some computable function. A sequence $X \in \Sigma^\omega$ is *$T(n)$-random* iff no martingale computable in $O(T(n))$-time succeeds on it.

It does not take that much computational power to have good randomness properties. For instance,

Proposition 1.8.6 ([Sch71])**.** *If $X \in \Sigma^\omega$ is n^2-random, then X satisfies the law of large numbers,*

$$\lim_n \frac{|\{i < n \mid X(i) = b\}|}{n} = \frac{1}{|\Sigma|} \text{ for all } b \in \Sigma.$$

And we can compute them efficiently,

Theorem 1.8.7 (See e.g. [FN15])**.** *Given a program for the time function T, one can compute a $T(n)$-random sequence in time $O(T(n) \cdot \log(T(n)) \cdot n^3)$-time.*

In Chapter 3 we will use the following property of $T(n)$-randomness.

Proposition 1.8.8. *Let $X \in \Sigma^\omega$, Γ be a non-trivial subset of Σ, and $g : \mathbb{N} \to \{0, 1\}$ be a computable function such that:*

1. *exists n_0 such that for all $n \geq n_0$, if $g(n) = 1$ then $X(n) \in \Gamma$, and*
2. *for infinitely many n, $g(n) = 1$.*

Then X is not computably random. Furthermore, if g is computable in $O(T(n))$-time with $T(n) = \Omega(n^2)$, then X is not $T(n)$-random.

Proof. The betting strategy is quite simple: wait until $n \geq n_0$ to start betting; then just bet \$1 to the symbols in Γ every time $g(n) = 1$. More formally, let $M : \Sigma \to \mathbb{Q}_2^{\geq 0}$ be defined as

$$M(\epsilon) = |\Gamma|$$
$$M(\sigma b) = \begin{cases} M(\sigma) - |\Gamma| + |\Sigma| & \text{if } M(\sigma) \geq |\Gamma|, |\sigma| \geq n_0, g(|\sigma|+1) = 1 \text{ and } b \in \Gamma \\ M(\sigma) - |\Gamma| & \text{if } M(\sigma) \geq |\Gamma|, |\sigma| \geq n_0, g(|\sigma|+1) = 1 \text{ and } b \notin \Gamma \\ M(\sigma) & \text{otherwise} \end{cases}$$

Now we have to see that M is a martingale computable in $O(T(n))$-time that succeeds on X.

1. If $M(\sigma) < |\Gamma|$ or $|\sigma| \leq n_0$ or $g(|\sigma|) = 0$, then $\sum_b M(\sigma b) = |\Sigma| M(\sigma)$. Else,

$$\begin{aligned} \sum_b M(\sigma b) &= \sum_{b \in \Gamma} M(\sigma b) + \sum_{b \notin \Gamma} M(\sigma b) \\ &= |\Gamma|(M(\sigma) - |\Gamma| + |\Sigma|) + (|\Sigma| - |\Gamma|)(M(\sigma) - |\Gamma|) \\ &= |\Sigma| M(\sigma). \end{aligned}$$

   Therefore, M satisfies the fairness condition (1.11) and, hence, is a martingale.

2. Let us study the complexity of computing M with the straightforward algorithm. If we denote with $F(n)$ the number of operations (in the worst case) on inputs of length n, then

$$F(n) = \begin{cases} O(1) & n \leq n_0 \\ F(n-1) + O(T(\log n)) + O(n) & n > n_0 \end{cases}$$

   where the second term in the summation is for the possible evaluation of g and the third is for the possible additions. Then, solving the recurrence,

$$F(n) = O(1) + \sum_{i=n_0}^{n} (c \cdot T(\log i) + d \cdot i)$$

   which is $O(T(n))$ if $T(n) = \Omega(n^2)$.

3. Now, to see that M succeeds on X,

$$\limsup_n M(X \upharpoonright n) = |\Sigma| + (|\Sigma| - |\Gamma|) \lim_n |\{n_0 \leq i \leq n \mid g(i) = 1\}| \tag{1.12}$$
$$= \infty \tag{1.13}$$

   Where, (1.12) follows from the definition of M and from assumption 1, and (1.13) follows from assumption 2.

□

Example 7. K is not computably random. Let $f : \mathbb{N} \to \mathbb{N}$ be a computable function such that $K = f(\mathbb{N})$ (K is c.e.) and let $h : \mathbb{N} \to \mathbb{N}$ be defined as $h(n) = f\left(\min_t[(\forall_{i<n} f(t) > h(i)) \vee n = 0]\right)$. It is not hard to see that $A = h(\mathbb{N})$ is a computable subset of K. Then, it follows from Observation 1.8.8 with $g := \chi_A$ and $\Gamma = \{1\}$ that K is not computably random.

Although every computable sequence is not computably random, a simple diagonal argument shows that we do not have a *general* computable betting strategy to succeed on all of them, that is

Theorem 1.8.9. *There is no computable martingale M that succeeds on every computable sequence.*

Proof. By way of contradiction, suppose that M is a computable martingale that succeeds on every computable sequence. Let $Y \in \{0,1\}^\omega$ be defined as: $Y(0) = 0$ and $Y(n+1) = b$ iff $M((Y \upharpoonright n)b) \leq M((Y \upharpoonright n)\bar{b})$. Hence, given that $M(Y \upharpoonright n) \geq M(Y \upharpoonright (n+1))$ for all n, Y is a computable sequence such that $\limsup_n M(Y \upharpoonright n) \leq M(\epsilon)$. A contradiction. □

In Chapter 4 we will need a randomness notion for infinite sequences for which a general (universal) procedure separating the random from the non-random (and, amongst this, the uncomputable) exists. Martin-Löf randomness is one such notion.

1.9 Martin-Löf randomness

Let $X \in \Sigma^\omega$ be the output of repeated throwing of a fair $|\Sigma|$-faced dice (equivalently, let X be sampled uniformly at random from Σ^ω). We expect X to have infinitely many occurrences of every symbol in Σ. Furthermore, we expect it to satisfy the *law of large numbers*, that is

$$\lim_n(|\{i < n \mid X(i) = b\}|/n) = 1/|\Sigma| \text{ for all } b \in \Sigma. \tag{1.14}$$

Such kind of properties of infinite sequences, which hold with probability 1, are called *laws of randomness*. We expect X to satisfy every such law. In measure-theoretic terms, X should belong to every set of (Lebesgue) measure 1. In other words, it will be *atypical* for X to *fail* to satisfy some of this laws. In measure-theoretic terms, belonging to the complement of some measure 1 class. Therefore, one would like to define a sequence as *random* if it doesn't belong to any null measure set. However, $\{X\}$ has measure 0, and so this is a vacuous definition. Not *every* null measure set should define a *test* for non-randomness.

A Martin-Löf test [ML66] is the formalization of a statistical test, intended to capture infinite sequences with certain patterns or special features. This 'detection' of non-random sequences must be computably approximable, with incrementing levels of accuracy or significance. A test is a collection of sets V_m of possible prefixes of sequences that do not look random. As we increase m, the identification of non-randomness gets more and more fine-grained, leaving in the limit a null measure set of non-random sequences. More formally, recalling that for a set $V \subseteq \Sigma^*$, $[V]$ denotes the set of infinite sequences having a string in V as a prefix, we have that

Definition 1.9.1 (Martin-Löf test)**.** Let $g : \mathbb{N}^2 \to \Sigma^*$ be a total computable function. A *Martin-Löf test (ML-test)* is a sequence $(V_m)_{m\in\mathbb{N}}$ of c.e. sets $V_m \subseteq \Sigma^*$ such that $V_m = \{g(m,n) \mid n \in \mathbb{N}\}$ and $\lambda([V_m]) \leq |\Sigma|^{-m}$ for all m, with $\lambda(\cdot)$ the uniform measure over Σ^ω.

The *Martin-Löf random (ML-random)* sequences will be those not detectable by any possible ML-test.

Definition 1.9.2 (failing a test)**.** A sequence Y is said to *fail* a ML-test $(V_m)_{m\in\mathbb{N}}$ iff $Y \in \bigcap_m [V_m]$. We also say that $(V_m)_{m\in\mathbb{N}}$ *captures* Y. If Y doesn't fail $(V_m)_{m\in\mathbb{N}}$, we say it *passes* it.

Informally, if $Y \in [V_m]$ for some m then we reject the hypothesis that Y is random with significance level $|\Sigma|^{-m}$. Observe that if $Y \in [\{\sigma_1, \sigma_2, \dots\}]$ then, for large enough n, we have that all the infinite sequences extending $Y \upharpoonright n$ belong to $[\{\sigma_1, \dots, \sigma_n\}]$. This last expresion can be seen as the n-th approximation of $[\{\sigma_1, \sigma_2, \dots\}]$. Hence if $Y \in \bigcap_m [V_m]$, then for every m there is n such that any extension of $Y \upharpoonright n$ is included in the n-th approximation of $[V_m]$. Finally,

Definition 1.9.3 (Martin-Löf random)**.** A sequence X is *Martin-Löf random (ML-random)* if it passes every ML-test. We denote with MLR the set of all ML-random sequences.

Example 8 (Chaitin's omega)**.** Let $\mathbf{M}$ be a Turing machine such that, for all $s \in \{0,1\}^*$, if $s \in \operatorname{dom}\mathbf{M}$, then $s[0]\dots s[i] \notin \operatorname{dom}\mathbf{M}$ for all $i < |s| - 1$ (i.e. the domain of $\mathbf{M}$ is *prefix-free*). Then, if $\mathbf{M}$ is universal, the real number $\Omega_{\mathbf{M}} \in [0,1]$ (equivalently, the sequence) defined as

$$\Omega_{\mathbf{M}} := \sum_{p\in\operatorname{dom}\mathbf{M}} 2^{-|p|}$$

is ML-random [Cha75]. Chaitin call these numbers *halting probabilities* (recall that $s \in \operatorname{dom}\mathbf{M}$ iff $\mathbf{M}$ halts on input s).

As we hinted at the end of the preceding section,

Proposition 1.9.4 (Universal ML-test)**.** *There is a ML-test $(U_m)_{m\in\mathbb{N}}$ such that for all $X \in \Sigma^\omega$, $X \in$ MLR iff X passes $(U_m)_{m\in\mathbb{N}}$.*

This implies that in the complement of the null measure set $\bigcap_m [U_m]$ we have all the ML-random sequences and so

Corollary 1.9.5. *A sequence X sampled uniformly at random from Σ^ω is ML-random with proability 1.*

The well known inclusion relationships between the notions of randomness for sequences just defined are

$$\text{MLR} \subsetneq \text{CR} \subsetneq T(n)\text{-randomness}$$

1.9.1 Relativized ML-randomness

In computability theory, we say that an algorithmic notion relativizes when it can be extended to the model of *oracle Turing machines* (see e.g. [Soa99, Section III.1]). Informally, these are Turing machines which, during the computation, can make finitely many queries to some (in general, non-computable) infinite sequence (the oracle).

Example 9 (K is computable relative to $\Omega_{\mathbf{V}}$)**.** Let us see how to compute K when given $\Omega_{\mathbf{V}}$ as an oracle. Let $i \in \mathbb{N}$ and $p = 0^i 1 i$. One can determine whether $\mathbf{V}(p)$ halts (i.e. whether $i \in K$) from only the first $|p|$ bits of $\Omega_{\mathbf{V}}$. Let $n = |p|$ and $\Omega_{\mathbf{V}}^{(n)}$ be $\Omega_{\mathbf{V}}$ truncated to the first n bits. We have $\Omega_{\mathbf{V}}^{(n)} < \Omega_{\mathbf{V}} < \Omega_{\mathbf{V}}^{(n)} + 2^{-n}$. Define $\Omega_{\mathbf{V}}[t] := \sum_{\mathbf{V}_t(s)\downarrow, |s| \leq t} 2^{-|s|}$ and note that $\lim_t \Omega_{\mathbf{V}}[t] = \Omega_{\mathbf{V}}$ and $\Omega_{\mathbf{V}}[t]$ is computable from t. Finally, let $t = \min_{t' \leq n}[\Omega_{\mathbf{V}}[t'] > \Omega_{\mathbf{V}}^{(n)}]$ and note that $\mathbf{V}(p)$ halts if and only if it halts in t steps, otherwise $\Omega_{\mathbf{V}} \geq \Omega_{\mathbf{V}}[t] + 2^{-n} > \Omega_{\mathbf{V}}^{(n)} + 2^{-n}$ a contradiction.

The notion of ML-randomness relativizes by simply letting the Turing machines enumerating the sets V_m in the above Definition 1.9.1 have access to some oracle X (this will be indicated with a superscript in the notation). Given two infinite sequences X and Y, we say that Y is *ML-random relative to* X if there is no ML-test $(V_m^X)_{m \in \mathbb{N}}$ such that $Y \in \bigcap_m [V_m^X]$. It is easy to see that when X is a computable sequence, being ML-random relative to X is equivalent to being ML-random, and that no sequence is ML-random relative to itself.

An important result, useful for our applications in Chapter 4, is that there exists a *universal oracle ML-test* $(U_m^X)_{m \in \mathbb{N}}$ such that, for all X and Y, Y is ML-random relative to X iff $Y \notin \bigcap_m [U_m^X]$. Since $\lambda \bigcap_m [U_m^X] = 0$, this implies that the set of ML-random sequences relative to X has measure 1 for all X. In other words, for any X, the sequence of independent throws of a $|\Sigma|$-faced dice is ML-random relative to X with probability 1.

1.10 Communication complexity

Classical communication complexity theory, introduced by Andrew Yao in 1979 [Yao79], studies the communication requirements in the distributed computation of functions. More formally, given a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, it asks how many bits, in the worst case, have to be exchanged between Alice holding input $x \in \mathcal{X}$ and Bob holding input $y \in \mathcal{Y}$ in order for him to output $f(x, y) \in \mathcal{Z}$ (see Fig. 1.2 for a schematic description).

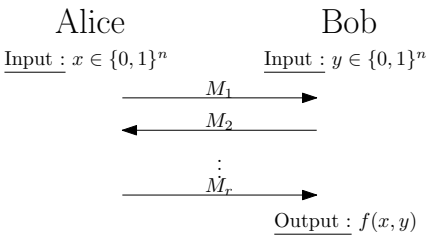

Fig. 1.2: Illustration of Communication Complexity's model

1.10.1 Communication complexity measures for computing functions

There are many variants of communication complexity. Here are some of the most well-known ones. We will first consider the standard deterministic model before extending it to the randomized model where the players can use randomness to be more efficient, and finally we will discuss how using quantum communication instead of classical can be even more effective.

Deterministic communication complexity

Let us first define what is called deterministic communication complexity, the most standard model which formalizes the definition given above. A deterministic communication protocol is an interactive protocol where at each step the player who speaks sends a bit to the other player, as a function of his input and the previous messages. We can formalize it using a walk on a tree where each node encodes the transcript of all the previous messages of the protocol, and depending on the current message, the players decide to move to one or the other of its children.

Definition 1.10.1 (deterministic communication protocol)**.** Let $f : \mathcal{I} \subseteq \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. A deterministic communication protocol Π which computes f is a binary tree where each internal node v is labelled either by a function $a_v : \mathcal{X} \to \{0,1\}$ or by a function $b_v : \mathcal{Y} \to \{0,1\}$ and each leaf is labeled by an output value (an element of $\mathcal{Z}$). On input (x, y), the execution of the protocol consist of traversing the tree as follows: when the node v belongs to Alice, she computes the value $a_v(x)$ and she sends it to Bob; when the node v belongs to Bob, he computes the value $b_v(y)$ and sends it to Alice. If the value is zero then they both move to the right child of v and to the left one otherwise. When they reach a leaf they output its value. We say that the protocol computes f if for each $(x, y) \in \mathcal{I}$ the execution of the protocol on (x, y) outputs $f(x, y)$. We consider that the last bit of the protocol is the output. The communication cost of a protocol Π is the height of the tree, we denote it by $CC(\Pi)$.

The deterministic communication complexity of a function f is the cost of the best deterministic protocol which computes f.

Definition 1.10.2 (deterministic communication complexity)**.** The deterministic communication complexity of f, denoted by $D(f)$ is defined by

$$D(f) = \min\{CC(\Pi) \mid \Pi \text{ computes } f\}.$$

To relax this problem, one can also define a notion of deterministic protocol for f which admits small error. Intuitively this can significantly decrease the complexity as a function can have a large complexity because few inputs are very difficult to handle, whereas on most of them it is easy to compute the value of the function.

Definition 1.10.3 (distributional communication complexity)**.** Let μ be a distribution on the input space $\mathcal{X} \times \mathcal{Y}$. We say that Π computes f with error ϵ according to μ if the probability over $(x, y) \sim \mu$ that the protocol outputs $f(x, y)$ on (x, y) is at least $1 - \epsilon$. The distributional communication complexity of f according to μ (denoted by $D_\mu(f)$)

is the cost of the best deterministic protocol that computes f with error ϵ according to μ.

$$D^{\epsilon}_{\mu}(f) = \min\{CC(\Pi) \mid \Pi \text{ computes } f \text{ with error } \epsilon \text{ according to } \mu\}.$$

Randomized communication complexity

Another way of considering communication complexity with error is to allow the players to act in a randomized fashion.

Definition 1.10.4 (public-coin communication protocol)**.** A public-coin communication protocol Π^{pub} is a distribution over deterministic protocols, run by first using shared randomness to sample an index r and then running the deterministic protocol Π_r. We say that Π^{pub} computes $f : \mathcal{I} \subseteq \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ with error ϵ if for each input $(x, y) \in \mathcal{I}$, the probability of choosing a protocol Π_r which outputs $f(x, y)$ is at least $1 - \epsilon$. The communication cost $R(\Pi^{pub})$ of such a protocol is the maximum number of bits that can be transmitted in any run of the protocol.

We can now define the notion of randomized communication complexity ($R_{\epsilon}(f)$) as the cost of the best public-coin communication protocol computing f with error ϵ .

Definition 1.10.5 (randomized communication complexity)**.**

$$R_{\epsilon}(f) = \min\{R(\Pi^{pub} \mid \Pi^{pub} \text{ computes } f \text{ with error } \epsilon\}.$$

The two notions of communication complexity with error (distributional and randomized) are closely related by Yao's famous Min-Max theorem [Yao83].

Theorem 1.10.6 ([Yao83])**.** $R_{\epsilon}(f) = \max_{\mu} D^{\epsilon}_{\mu}(f)$.

Quantum communication complexity

Fifteen years after he defined the classical model for communication complexity, Andrew Yao [Yao93] proposed to study the model where, instead of sending bits, the players can also exchange qubits. The main question was to understand whether this quantum model could be stronger than the classical one, i.e., if there is some function for which we save a lot of communication using quantum communication. Despite Holevo's theorem [Hol73] which states than no more than n bits of classical information can be communicated with n qubits (if the players are not allowed to use entanglement), some separations have been proven between the classical and the quantum models. The first significant separation is due to Buhrman et al. [BCW98]. They proved, defining some variant of the equality function, that quantum communication can be exponentially better than classical communication in the zero-error case. Then Ran Raz [Raz99a] proved an exponential separation for the bounded-error case. Formally, in a quantum communication protocol, each player has a working space on which, at each step, he applies some unitaries depending on his input and then sends some part of his space (the message) to the other player. At the end of the protocol, one player does a measurement to determine the output of the protocol.

Definition 1.10.7 (quantum communication protocol)**.** A quantum communication protocol Π^{qubit} is defined as the follows: Alice has an input x and Bob y. Their working space is a quantum state separated into three work spaces (three registers): one private register for Alice, one private register for Bob and one register for the communication. At the beginning of the protocol, the state is $|0\rangle_A|0\rangle_M|0\rangle_B$. At each step of the protocol, if it is Alice's turn to talk (let say on the k-th round), she applies some unitary $U_A^{k,x}$ which is a function of her input x on her register and the communication register. This operation corresponds both to her local computation and to putting a message in the communication channel. Here the size of her quantum message corresponds to the number of qubits of the communication register which has been changed by her operation. Since in this round nothing happens to Bob's register, the operation on the overall quantum state is $U_A^{k,x} \otimes \mathbb{1}_B$ where $\mathbb{1}_B$ is the identity on Bob's register. When we are in round k where it is Bob's turn to talk, he applies one unitary operation $U_B^{k,y}$ on his register and the communication register, so he applies $\mathbb{1}_A \otimes U_B^{k,y}$ to the current quantum state. After t steps of this protocol the quantum state is $(U_A^{t,x} \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes U_B^{t-1,y}) \dots (\mathbb{1}_A \otimes U_B^{2,y})(U_A^{1,x} \otimes \mathbb{1}_B)|0\rangle_A|0\rangle_M|0\rangle_B$. At the end of the protocol Bob applies some measurement on his register and the output of the protocol is the value he observes. We say that such a protocol computes f with error at most ϵ if for each $(x, y) \in \mathcal{I}$ the probability that it outputs $f(x, y)$ is at least $1 - \epsilon$. The cost $QC(\Pi^{qubit})$ of such a protocol is the maximum number of qubits sent over all the possible inputs where at each step the number of qubits sent corresponds to the number of qubits which have changed in the communication register.

We can now define the quantum communication complexity of f (denoted by $Q_\epsilon(f)$) as the cost of the best quantum protocol which computes f with error ϵ.

Definition 1.10.8 (quantum communication complexity)**.**

$$Q_\epsilon(f) = \min\{QC(\Pi^{qubit}) \mid \Pi^{qubit} \text{ computes } f \text{ with error } \epsilon\}.$$

1.10.2 Communication complexity measures for distributions

We will study here the communication complexity of simulating distributions. In this model Alice has x, Bob has y and this input defines a distribution on some output space $\mathcal{A} \times \mathcal{B}$ denoted by $p(..|x, y)$. Using communication and randomness, their goal is to respectively output a and b such that the probability of outputting (a, b) on (x, y) is $p(a, b|x, y)$. We use the following notation for communication complexity of distributions. $R_\epsilon(p)$ (resp. $Q_\epsilon(p)$) is the minimum amount of classical communication (resp. quantum) necessary to reproduce the distribution p in the worst case, up to ϵ in total variation distance for all x, y. We write $|\mathbf{p} - \mathbf{p}'| \leq \epsilon$ to mean that for any x, y, $\sum_{a,b} |p(a, b|x, y) - p'(a, b|x, y)| \leq \epsilon$.

Boolean function as a special case of distribution

Boolean (and other) functions can be cast as a sampling problem as follows. Consider a Boolean function $f : \mathcal{I} \subseteq \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ whose communication complexity we wish to study (non-Boolean functions and relations can be handled similarly). First,

we split the output so that if $f(x, y) = 0$, Alice and Bob are required to output the same bit, and if $f(x, y) = 1$, they output different bits. Let us further require Alice's marginal distribution to be uniform, likewise for Bob, so that this sampling problem defines a distribution. Call the resulting distribution $\mathbf{p}_f$. If $\mathbf{p}_f$ were local (Definition 1.5.2), f could be computed with one bit of communication using shared randomness: Alice sends her output to Bob, and Bob XORs it with his output. More precisely, from a boolean function we can construct the following distribution:

Definition 1.10.9. Let $f : \mathcal{I} \subseteq \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, define $p_f(a, b|x, y) = \frac{1}{2}$ if $f(x, y) = a \oplus b$ and 0 otherwise, where a and b are booleans.

Theorem 1.10.10. $R_\epsilon(\mathbf{p}_f) \leq R_\epsilon(f) \leq R_(\mathbf{p}_f) + 1$.

Proof. To compute $f(x, y)$ from a protocol for $\mathbf{p}_f$, Alice sends a to Bob who can evaluate $a \oplus b$ which is the value of $f(x, y)$. To simulate $\mathbf{p}_f$ from a protocol computing f Alice and Bob use public randomness to pick a random bit a. When the protocol for f outputs z, Alice outputs a and Bob outputs $z \oplus a$. The relation still holds with error ϵ. □

We can notice that for any Boolean function f, the distribution $\mathbf{p}_f$ described is non-signaling (Definition 1.5.10) since the marginals are uniform.

RESUMEN DEL CAPÍTULO

En este capitulo se presentan los objetivos de la tesis, se resumen los principales resultados obtenidos y se introducen los conceptos y antecedentes necesarios para su entendimiento.

La computación cuántica y la información cuántica [NC11] tratan acerca de utilizar el andamiaje teórico de la mecánica cuántica para diseñar protocolos y/o construir sistemas que realicen tareas de procesamiento de la información que son clásicamente difíciles o aun imposibles de realizar. La aplicación de los principios mecánico-cuánticos al estudio de tareas de procesamiento de información ha producido significativos avances tales como: el algoritmo de Shor para factorizar enteros de manera eficiente [Sho97], protocolos de distribución de claves que basan su seguridad en las leyes de la física en lugar de en la conjeturada intractabilidad de un problema matemático [BB84, Eke91] y protocolos de amplificación de aleatoriedad [CK11, CR12, GMDLT$^+$13], por nombrar sólo algunos.

El objetivo de esta tesis es analizar algunos de los conceptos fundamentales de la mecánica cuántica, fundamentales para los avances antes mencionados, desde una perspectiva teórico-computacional. El enfoque va a ser doble: por un lado, vamos a analizar las hipótesis detrás de las resultados; por el otro, vamos a asumir la validez de las hipótesis para luego cuantificar la ganancia computacional que resulta de aplicar-los los consecuentes principios mecánico-cuánticos. Específicamente, en el lado de las hipotesis, vamos a recurrir a las teorías de la aleatoriedad algorítmica y la inferencia inductiva para estudiar el impacto de reemplazar la hipótesis de aleatoriedad por el uso de pseudoaleatoriedad el algunas partes de la teoría y la práctica. Luego, en el lado de las aplicaciones, vamos a estudiar la ventaja que ofrece la cuántica en el área de complejidad computacional.

A continuación se enumeran las preguntas principales junto con las respuestas obtenidas:

- Pregunta: En un experimento de Bell, ¿alcanza con que la elección de las mediciones sea pseudoaleatoria para poder concluir la no-localidad de las correlaciones observadas a partir de la violación de una desigualdad de Bell?

- Respuesta: Si en un experimento de Bell bipartito las entradas de al menos una de las partes son pseudoaleatoreas entonces, si se conoce una cota superior computable a la complejidad temporal de la función usada, hay un modelo local para explicar cualquier violación de Bell.

- Pregunta: ¿Tiene alguna consecuencia obeservacional usar pseudoaleatoriedad en lugar de aleatoriedad la preparación de estados mixtos (como hacen en, por ejemplo, [AB09a, LKPR10])?

- Respuesta: Hay un protocolo para distinguir cualquier mezcla pseudoaleatoria de estados cuánticos puros del estado máximamente mixto.

- Pregunta: ¿Es compatible con nuestras teorías físicas actuales una Naturaleza en la cual la salida de experimentos de no-localidad es producida de manera computable y suplementada con algún tipo de señalización escondida?

- Respuesta: Una teoría de este tipo está en contradicción con la relatividad especial pues damos un protocolo para usar cajas no-locales que se comporten de tal forma para comunicar información entre puntos distantes de manera instantánea.

- Pregunta: ¿Es la no-localidad la responsable de la ventaja cuántica en comunicación computacional?

- Respuesta: Para un gran familia de funciones, mostramos cómo construir desigualdades de Bell y distribuciones cuánticas que las violan en una magnitud que es exponencial en la diferencia entre sus complejidades comunicacionales clásicas y cuánticas.

2. THE COMPUTABILITY LOOPHOLE

Bell proved that the correlations predicted by quantum mechanics between the outputs of suitably chosen local measurements over parts of an entangled quantum system cannot be accounted for by any local hidden-variable model in which the hidden-variables are independent of the local measurement choices. Recent theoretical work has demonstrated that models that relax this measurement independence assumption, allowing for a modest correlation between the joint measurement settings and any causal influence on the measurement outcomes, can reproduce the quantum correlations [BSHC85, Hal10, BG11, TSS13, PRB$^+$14]. The approach to the question of measurement independence has so far been of a statistical nature. That is, people have studied different measures of correlation or dependence between the hidden-variables and the inputs random variables and provided quantitative lower bounds necessary for the observation of non-locality on different Bell scenarios. In this chapter we switch from a *process* approach to a *product* approach and study the problem from an algorithmic information perspective. Specifically, we study what happens with the validity of concluding the non-locality of the observed correlations in a Bell experiment from the violation of a Bell inequality when the sequences of inputs are computable (e.g. the output of private pseudorandom number generators (PRNGs)). We show that in such a setting an eavesdropper without access to the PRNGs can prepare local boxes that seem non-local provided she knows an upper bound on the time computational complexity of the pseudorandom functions and has access to the inputs and outputs of previous rounds (**Theorem 2.2.1**). In other words, we show that the sequences of inputs to a Bell test being computable opens up a loophole, which we call: *the computability loophole* [BdlTS$^+$16].

2.1 Measurement independence

We have characterized the local distributions $\mathcal{L}$ as those admitting a local deterministic λ-independent hidden-variable model (Proposition 1.5.7). The property of λ-independence [BY08], stating the independence between the measurement choices and the hidden variables, has received various other names in the literature: *measurement independence* [Hal10], *free will* [CR12] and *no-conspiracy* [Nor11]. Trivially, any distribution $\mathbf{p}$ can be reproduced by a local model in which the hidden-variables are fully dependent on the inputs. Furthermore, this still holds (for non-signaling distributions) if the dependence is with only one of the parties' inputs.

Proposition 2.1.1. *For every non-signaling distribution* $\mathbf{p}$ *there is a local model with* $q(\lambda|y) = q(\lambda)$.

Proof. First, notice that for any non-signaling distribution $\mathbf{p} \in \mathcal{C}$ we have

$$\begin{aligned} p(a,b|x,y) &= p(a|x,y)p(b|a,x,y) \\ &= p(a|x)p(b|a,x,y) \qquad (\mathbf{p} \in \mathcal{C}) \end{aligned}$$

Let $X \in \mathcal{X}^\omega$ be the sequence of Alice's inputs. The local model is as follows. At round n, the hidden variable λ is a vector $[a_n, X(n)]$, with a_n sampled from $p(a|X(n))$; Alice's box outputs a_n and Bob's box, on input y, (locally) outputs a sample from $p(b|a_n, X(n), y)$.

Now,

$$\begin{aligned} p(a,b|x,y) &= p(\lambda = [a,x]|x)p(b|\lambda = [a,x], y) \\ &= p(\lambda[0] = a|x)p(\lambda[1] = x|x)p(b|a,x,y) \\ &= p(a|x)p(b|a,x,y) \end{aligned}$$

with the last equality following from $p(\lambda[1] = x|x) = 1$. □

However, complete independence of the measurement choices from any physical parameter influencing the measurement outputs, which is typically justified by an appeal to the experimentalist's free will [BSHC85], may seem too strong of an assumption and one may wonder if we can still observe genuine non-locality when this requirement is, somehow, relaxed. This is of special importance in cryptographical uses of Bell's theorem, where the choice of measurement is delegated to physical systems whose random behaviour cannot be guaranteed [GMDLT$^+$13, MPA11].

In order to study the possibility of relaxing the λ-independence assumption while still being able to separate local from non-local models, we need a way of quantifying the dependence between the measurement choices and the hidden variables. In the following, we review the measures proposed in the literature and the quantitative results they allow to draw from Bell violations.

In [Hal10], it is considered

$$M := \sup_{x,x',y,y'} \int d\lambda |p(\lambda|x,y) - p(\lambda|x',y')|, \tag{2.1}$$

the 'maximum distance' between the distributions of the underlying variable for any two pairs of measurement settings. Clearly, with $M = 0$ we have full measurement independence and with $M = 2$ we have that there are at least two particular pairs of inputs (x,y) and (x',y') such that for any λ at most one of these settings is possible and hence there is no free will in deciding between them. With this, the fraction of measurement independence can be quantified via

$$F := 1 - M/2. \tag{2.2}$$

It is then shown in [Hal10] that only by giving up 14% of measurement independence (i.e. with $F \leq 86\%$), there is a local deterministic model of the singlet correlations.

Another natural way of quantifying the dependence between the hidden variables and the measurement choices is through their mutual information [BG11]

$$I(x, y : \lambda) = H(x, y) + H(\lambda) - H(x, y, \lambda), \tag{2.3}$$

where H is the Shannon entropy. When x and y are independent of λ, $I(x, y : \lambda) = 0$. On the other hand, if x and y are functions of λ, then $I(x, y : \lambda) = H(x, y)$. In [BG11] it is shown that for any Bell experiment with two inputs per party, as in the CHSH scenario, there is a local model accounting for the observed correlations with an amount of mutual information not bigger than one. For example, in the ideal case of $H(x, y) = 2$, the model has $I(x, y : \lambda) \approx 0.85$.

In the next section we take a more operational approach to the question of measurement independence. We consider the case in which the parties in a Bell test use pseudorandom numbers generators (PRNGs) to choose the inputs. We show that, in such scenario, there is a local model in which the hidden variables, after finitely many rounds and without access to the PRNGs used by the parties, start to perfectly predict the future measurement choices, allowing them to fake any non-local behaviour. Granted, the possibility that physical hidden variables behave in this way is quite implausible and conspiratorial (a feature, albeit, shared with most of the other local models which exploit experimental loopholes [Hal10, CH74, LG04]). However, the scenario becomes significantly plausible when we consider a cryptographical context in which we are not testing quantum mechanics but rather using devices as black boxes received from some untrusted provider and basing the security of our protocols on the observed non-local behaviour.

2.2 The loophole

It is convenient for what follows to rephrase the standard Bell scenario in cryptographic terms, as in [BCH$^+$02, PAM$^+$10, PM13]. In this approach, Alice and Bob get their boxes from a non-trusted provider Eve. This cryptographic approach to Bell tests, we believe, makes it easier to understand the implications of our results for device-independent protocols based on non-locality. Nevertheless, our results also apply to the standard context in which Bell inequalities are used, namely, to test the possible existence of a local model explaining quantum correlations [EPR35, Bel64]. There, the local model can be seen as the eavesdropper that tries to reproduce the observed correlations, possibly by exploiting loopholes in the implementation.

We will consider an scenario in which Eve, in preparing the boxes for round $n + 1$ of the experiment, have access to all inputs and outputs of the previous n rounds; or, equivalently, that the boxes she prepares can communicate the inputs used in the previous rounds as shown in Figure 2.1 [BCH$^+$02, PAM$^+$10, PM13]. In this *two-sided memory* scenario [BCH$^+$02], the local distributions are those which can be written as

$$P(a^{(n)}, b^{(n)}|x^{(n)}, y^{(n)}) = \sum_{\lambda} p(\lambda)p(a^{(n)}|x^{(n)}, M, \lambda)p(b^{(n)}|y^{(n)}, M, \lambda)$$

where $M = [x_0, \ldots, x_{n-1}, y_0, \ldots, y_{n-1}, a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}]$. That is, the local distributions as those for which the outputs at round $n + 1$ are generated from the local

inputs plus the information of the previous rounds M and some shared randomness λ. It is shown in [BCH$^+$02] that letting the eavesdropper (equivalently, the boxes she generates) have *memory* of the previous inputs and outputs still allows one to see genuine non-locality. In this chapter we show that this is no longer the case if, in addition to the devices having memory, we let Alice or Bob choose their inputs pseudorandomly. The main intuition is that if Eve, or more precisely the devices she prepares, are able at some point to learn the algorithm generating the inputs, she could use this information to produce a fake Bell violation.

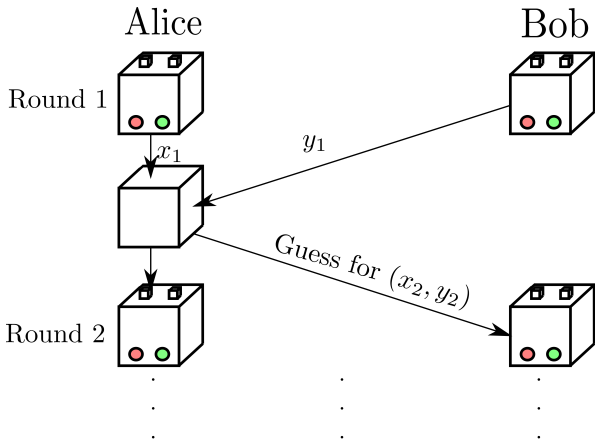

Fig. 2.1: Scheme for the Bell inequality computability loophole. After each round i, Alice's box receives Bob's last choice of measurement y_i. Using all previous choices of inputs for both parties, Alice's box makes a prediction for what the inputs of the next round will be.

Let us assume that one party, say Alice without loss of generality, uses an algorithm to choose her inputs. In formal terms, this means that there is a computable function $f_A : \mathbb{N} \to \{0, 1\}$ such that $f_A(i)$ tells Alice to press the left (0) or the right (1) button at the i-th round. As we saw in Proposition 2.1.1, for every target non-signaling distribution $\mathbf{p}$ and any function f_A giving the inputs of one of the parties, there is a local model. In other words, for every non-signaling $\mathbf{p}$, if Eve knows f_A, there is a strategy for her (dependent on f_A) to prepare local boxes generating $\mathbf{p}$.

We would like to have a strategy for Eve that works independently of f_A. We will, however, assume the following further hypothesis: Eve knows *some* computable function $T : \mathbb{N} \to \mathbb{N}$ which upper bounds the running time needed to compute f_A. For instance, Eve knows that f_A is computable in $O(T(n))$-time for, say, $T(n) = 2^{2^n}$ —though the algorithm that Alice is actually running may take, say, $O(n^2)$.

Now, our main result, the existence of a loophole (i.e. a strategy for Eve to fake any target non-signaling distribution $\mathbf{p}$), follows from the fact that the class of functions computable in $O(T(n))$-time is identifiable by next value (Corollary 1.7.11). Formally, our result states that,

Theorem 2.2.1. *For every computable function $T : \mathbb{N} \to \mathbb{N}$ and every non-signaling distribution $\mathbf{p}$, there is a strategy for an eavesdropper Eve to prepare local boxes for every round of a Bell experiment by having memory of the inputs and outputs in previous*

rounds, reproducing $\mathbf{p}$ *whenever Alice's inputs are given by a function* $f_A : \mathbb{N} \to \mathbb{N}$ *computable in* $O(T(n))$*-time.*

Proof. The strategy for Eve is an adaptation of the proof of Proposition 2.1.1 to our scenario with memory. Let

$$M = [x_0, \ldots, x_{n-1}, y_0, \ldots, y_{n-1}, a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}]$$

be the inputs and outputs of the first n rounds of the Bell test and, by Corollary 1.7.11, let $g_T : \mathbb{N} \to \mathbb{N}$ be a predictor for the class of functions computable in $O(T(n))$-time. Note that $[\langle 0, x_0\rangle, \ldots, \langle n-1, x_{n-1}\rangle] = f_A^{n-1}$. On the n-th round, Eve will, first, sample a string $\lambda \in \mathcal{A}^{|\mathcal{X}|}$ from the product distribution $\prod_{x\in\mathcal{X}} p(\cdot|x)$. Then, she will prepare and distribute boxes A_n to Alice and B_n to Bob such that

- A_n, on inputs x_n and M, outputs $\lambda[g_T(f_A^{n-1})]$ (disregarding x_n) and
- B_n, on inputs y_n and M, outputs b_n locally sampled from $p(b|\lambda[g_T(f_A^{n-1})], g_T(f_A^{n-1}), y_n)$.

The fact that the strategy reproduces $\mathbf{p}$ now follows from the analysis in the proof of Proposition 2.1.1 and from the fact that, by assumption, there exists a (number of round) n_f such that for all $n \geq n_f$, $g_T(f_A^{n-1}) = f_A(n) = x_n$ (note that the *finite* contribution to the statistics of the rounds before the n_f-th is negligible). □

At this point, we can further clarify the need to assume a bound on the complexity of f_A. As we saw, the loophole is based on the ability to program a predictor (in the sense described above) for functions belonging to a given class. However, the class of all total computable functions is not predictable (c.f. Theorem 1.7.4). We could have chosen other ways to restrict the class of functions, but computational resources seemed the more natural.

Despite being necessary for the protocol, one can justify the time complexity assumption on the following grounds:

1. it is natural to require that the time Alice and Bob take to choose their measurements on each round is bounded, and
2. the number of computational steps per second that a physical system of mass m can perform is upper bounded by $2(mc^2)/\pi\hbar$ [Llo00].

These two facts imply that the number of computational steps that Alice's and Bob's algorithms can take on each round n is bounded by a constant and hence, their computational complexity is, at most, linear in n (and, so, exponential in $|n|$, the size of n).

Regarding the complexity of Eve's protocol, there are two measures that one can study. First, there is time complexity of the predictor g_T: if $T(n)$ is the upper bound assumed by Eve for the running time of Alice's algorithm, then $g_T \in O(T(n)\cdot\log(T(n))$ (for $T(n)$ time constructible, see [AB09b, §1.3]). Second, there is the number M of mistakes that g_T will make before starting to guess correctly. Using the *halving algorithm* of Barzdin and Freivalds (see [ZZ08, Thm. 6]), the learning process can be carried out

in such a way that $M \leq O(\max(l, \log(c)))$, where l is the length of Alice's algorithm and c is such that it runs in time $c \cdot t(n)$.

This means that Eve will not require too many rounds, in terms of l and c, to fake non-locality. That is, if we look at the distribution generated in the first n rounds, the fraction of inputs-outputs that will not serve Eve's purpose of faking a non-local distribution is upper bounded by M/n, which vanishes with increasing n. Therefore, if Alice wants to make this number of rounds large, then she either has to use a very long program or an enormous time constant.

2.3 Discussion

In this chapter we showed that if either Alice or Bob choose the inputs for a Bell experiment in a computable way, an eavesdropper able to bound their time computational complexity can prepare deterministic devices and make them believe they have non-local boxes, thus creating a loophole. For the loophole to apply, the boxes should communicate between rounds and adapt accordingly, as for instance studied in the context of the memory loophole [PM13, PAM$^+$10]. There is no way of preventing this form of communication, unless some assumptions regarding the shielding of the devices are enforced, or by imposing that all the measurements in the Bell test by one of the parties are space-like separated from those by the other party.

It is relevant to place these considerations in the context of recent "loophole-free" Bell experiments [HBD$^+$15, GVW$^+$15, SMSC$^+$15]. In all these experiments the choice of measurements was performed using the fast quantum random number generator (QRNG) of [AAM$^+$15]. Thus, assuming the validity of quantum physics, these experiments are free from the computability loophole introduced here. However, one may argue that it is rather undesirable, and even circular, to depend on the validity of a non-local theory, such as quantum physics, to test non-locality. The use of random numbers of quantum origin is better justified in device-independent protocols based on non-locality, as the validity of quantum physics is assumed for many of them.

RESUMEN DEL CAPÍTULO

Bell demostró que las correlaciones predecidas por la mecánica cuántica entre las salidas de mediciones locales apropiadamente elegidas sobre partes de un sistema cuántico entrelazado no pueden ser explicadas por un modelo local de variables escondidas en el cual las variables escondidas son independientes de la las elecciones locales de medición. Trabajo teórico reciente ha demostrado que modelos que relajan esta suposición de independencia de mediciones, permitiendo por una correlación modesta entre las configuraciones de mediciones conjuntas y cualquier influencia causal sobre los resultados de la medición, pueden reproducir las correlaciones cuánticas[BSHC85, Hal10, BG11, TSS13, PRB$^+$14]. Por ahora, el enfoque a la cuestión de independencia de mediciones ha sido de una naturaleza estadística. Esto es, se han estudiado diferentes medidas de correlación o dependencia entre las variables escondidas y las variables aleatorias de entrada y se han proveído cotas inferiores cuantitativas necesarias para la observación de no-localidad en diferentes escenarios de Bell.

En este capítulo cambiamos de un enfoque de *proceso* a un enfoque de *producto* y estudiamos el problema desde una perspectiva de información algorítmica. Específicamente, estudiamos qué pasa con la validez de concluir la no-localidad de las correlaciones observadas en un experimento de Bell a partir de la violación de una desigualdad de Bell cuando las secuencias de las entradas son computables (e.g. la salida de generadores privados de números pseudoaleatorios (PRNGs, por sus siglas en ingles)). Mostramos que en tal contexto, un espía sin acceso al PRNGs puede preparar cajas locales que parecen no-locales asumiendo que conoce una cota superior en la complejidad computacional de tiempo de las funciones pseudoaleatorias y tiene acceso a las entradas y salidas de las rondas anteriores (**Theorem 2.2.1**). En otras palabras, mostramos que las secuencias de entradas para un test de Bell siendo computable abre un loophole, el cual llamamos: *el loophole de computabilidad* [BdlTS$^+$16].

Es relevante ubicar este resultado en el contexto de los recientes experimentos de Bell "loophole-free" [HBD$^+$15, GVW$^+$15, SMSC$^+$15]. En todos estos experimentos, la elección de qué medición realizar se hizo utilizando el generador cuántico de números aleatorios de [AAM$^+$15]. Por lo tanto, asumiendo la validez de la física cuántica, estos experimentos están libres del loophole de computabilidad. Sin embargo, uno podría argumentar que es un tanto indeseable, y quizás aún circular, depender de la validez the una teoría no-local, como es la mecánica cuántica, para testear no-localidad. El uso de números aleatorios de origen cuántico está mejor justificado en los llamados "protocolos independientes-del-dispositivo" basados en no-localidad, dado que en mucho de ellos se asume la validez de la teoría cuántica.

3. COMPUTABLE NON-LOCALITY ALLOWS FOR FASTER THAN LIGHT SIGNALING

It is a consequence of Bell's theorem [Bel64] that any deterministic hidden-variable account of the non-local correlations that quantum mechanics predicts and which we are now almost certain [HBD$^+$15, GVW$^+$15, SMSC$^+$15] that Nature exhibits, must allow for the existence of some kind of signaling mechanism that links distant measurement choices and outcomes. But, since quantum correlations are non-signaling, such signaling mechanism must be restricted to the hidden-variables and not reach the phenomenological level.

Some examples of deterministic accounts of non-local correlations are: the hidden variable model with communication of Toner and Bacon [TB03b] and, more prominently, Bohmian mechanics [Boh52]. For those models that use classical communication to mimic non-locality, one can in fact study the amount of communication needed (see, for example, [RT09, SZ08, DKLR11]). In all these theories, although the outputs at each round of a Bell test are *determined* given the inputs and the hidden-variable, the particular hidden-variable is sampled from some (non-deterministic) distribution. In this chapter we study the class of deterministic models for non-local correlations in which the hidden variables are not chosen "randomly" but pseudorandomly. In principle, the sequence of hidden variables for a given experiment is experimentally inaccessible; we wonder whether these sequences being computable has any observational consequences.

Our main result is to show that deterministic hidden-variable models of non-local correlations need to be uncomputable if we want to prevent those correlations from being signaling. In other words, we show that if the deterministic model is computable, the hidden-signaling mechanism used to exhibit non-locality can be extracted at the observational level and used for the communication of information between the parties provided a computable upper bound is known for the time computational complexity of the computable signaling (**Theorem 3.2.7**). More specifically, we give a protocol to perform one-way communication between two observers holding computable non-local boxes in some known time computational complexity class [BdlTS$^+$17].

There are a few previous results in this direction. First, our result has a flavour similar to [Yur00], where it is argued that the possibility to algorithmically compress the outputs of measurements over certain bipartite quantum states would allow for signaling. However, we obtain our result in a device-independent scenario, that is, without assuming quantum mechanics. Second, in [Wol15], computability of the outputs implying signaling is proven for the PR-box and for any non-local boxes violating the chained Bell inequality or winning any pseudotelepathy game (nonlocal games having a quantum strategy which wins with probability one). Our result is that this is true for

any non-local correlations. We provide an explicit communication protocol. Finally, the question of the computability of the sequences of outputs, but without relating it to the possibility of signaling, has also been studied for contextuality scenarios [ACCS12], through the localization of *value indefinite* observables [ACS15].

3.1 The scenario

We consider a standard bipartite Bell scenario, where we have players Alice and Bob with Alice holding a box having inputs in $\mathcal{X}$ and outputs in $\mathcal{A}$ and Bob holding a box having inputs in $\mathcal{Y}$ and outputs in $\mathcal{B}$. Our goal is to study deterministic and computable models that reproduce non-local correlations. This means that:

1. *Determinism.* The output of Alice's and Bob's boxes at each round n are functions $A_n : \mathcal{X} \times \mathcal{Y} \to \mathcal{A}, B_n : \mathcal{X} \times \mathcal{Y} \to \mathcal{B}$ of the inputs.
2. *Computability.* The mappings $n \mapsto A_n$ and $n \mapsto B_n$ are computable.
3. *Non-locality.* There exists a Bell functional B such that, when Alice and Bob choose their inputs uniformly at random, the expected value for B over the statistics collected up to round N violates the local bound as N goes to infinity.

Items 1 and 2 are formalized as follows: there are computable functions $A : \mathcal{X} \times \mathcal{Y} \times \mathbb{N} \to \mathcal{A}, B : \mathcal{X} \times \mathcal{Y} \times \mathbb{N} \to \mathcal{B}$, representing Alice's and Bob's boxes respectively, such that $A(x, y, n)$ [resp. $B(x, y, n)$] represents the output of the n-th round of Alice's [resp. Bob's] box when Alice's input is $x \in \mathcal{X}$ and Bob's input is $y \in \mathcal{Y}$. See Figure 3.1 for a schematic representation. On the other hand, item 3 is formalized through the following definition:

Definition 3.1.1 (non-local boxes)**.** A pair of boxes A, B is *non-local* if there exist a Bell inequality

$$\sum_{a,b,x,y} B_{a,b,x,y} p(a, b|x, y) \leq B_L$$

such that, if the inputs in a Bell experiment are chosen uniformly at random, then the sequences $Z \in \mathcal{A}^\omega$ and $W \in \mathcal{B}^\omega$ of outputs from A and B respectively are such that

$$\lim_N E(B_N) > B_L,$$

where $E(B_N)$ is the expected value of the random variable

$$B_N := \frac{|\mathcal{X} \times \mathcal{Y}|}{N} \sum_{a,b,x,y} B_{a,b,x,y} |\{i \leq N \mid X(i) = x, Y(i) = y, Z(i) = a, W(i) = b\}|.$$

Note that Definition 3.1.1 is general enough to cover the usual non-deterministic scenario as well.

As we said in the introduction, because we are looking at deterministic boxes generating non-local correlations, their outputs have to depend on each other's input. Since the boxes are computable, this is the only information they need to share, as any other

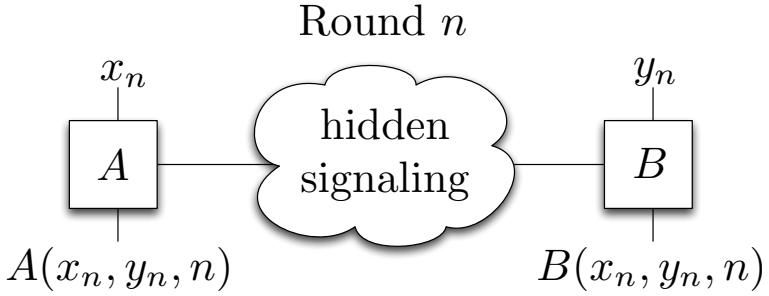

Fig. 3.1: Schematic representation of the scenario considered. Two distant observers, Alice and Bob, run a Bell test by implementing measurements on two systems. The observed correlations are described by a hidden-signaling mechanism plus computable functions determining the outputs given the inputs at each round n.

necessary data can be computed from the inputs. It is important to note that, although it seems that our toy model is signaling, and therefore it would not come as a surprise that Alice can signal to Bob, this is not the case. The model uses signaling for its internal workings but does not necessarily allow Alice and Bob to send information to each other. For instance, if one does not impose the computable condition to functions A and B, one can easily simulate quantum mechanics in a way that is completely equivalent and indistinguishable from standard quantum theory.

It is easy to see that, if the dependence between distant inputs and outputs happens in only finitely many rounds, the boxes are essentially local. Therefore, we have that:

Lemma 3.1.2. *If A and B are a pair of deterministic non-local boxes, then for infinitely many values of n, there exists $x \in \mathcal{X}$ and $y, y' \in \mathcal{Y}$ such that $A(x, y, n) \neq A(x, y', n)$ or there is $y \in \mathcal{Y}$ and $x, x' \in \mathcal{X}$ such that $B(x, y, n) \neq B(x', y, n)$.*

Proof. By way of contradiction, suppose that there exists n_0 such that for all $n \geq n_0$,

$$\forall x\ \forall y, y'\ A(x, y, n) = A(x, y', n) \text{ and}$$
$$\forall x, x'\ \forall y\ B(x, y, n) = B(x', y, n)$$

and consider the Bell inequality,

$$B = \sum_{a,b,x,y} B_{a,b,x,y} p(a, b|x, y) \leq B_L. \tag{3.1}$$

We will show that when the sequences of inputs X and Y are sampled uniformly and independently from $\mathcal{X}^\omega$ and $\mathcal{Y}^\omega$ respectively and the sequences of outputs are $Z(n) = A(X(n), Y(n), n)$ and $W(n) = A(X(n), Y(n), n)$, then $\lim_N E(B_N) \leq B_L$, thus contradicting the assumption that A and B are non-local boxes.

Define $A'(x, n) := A(x, x_0, n)$ and $B'(y, n) = B(0, y_0, n)$ for some fixed $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$. Then, for all $n \geq n_0$ and for all x and y, $A'(x, n) = A(x, y, n)$ and $B'(y, n) = B(x, y, n)$.

Let

$$B_N^n = \sum_{a,b,x,y} \delta_{X(n)=x} \delta_{Y(n)=y} \delta_{Z(n)=a} \delta_{W(n)=b}.$$

We have that,

$$\begin{aligned} E(B_N^n) &= E\left(\sum_{a,b,x,y} \delta_{X(n)=x}\delta_{Y(n)=y}\delta_{Z(n)=a}\delta_{W(n)=b}\right) \\ &= \sum_{a,b,x,y} E\left(\delta_{X(n)=x}\delta_{Y(n)=y}\delta_{Z(n)=a}\delta_{W(n)=b}\right) \\ &= \sum_{a,b,x,y} p^n(x,y)p^n(a,b|x,y) \end{aligned}$$

where $p^n(x,y)$ is the probability that the inputs in the n-th round are (x,y) and $p^n(a,b|x,y)$ is the probability that in the n-th round the outputs of the boxes are $(a,b) \in \mathcal{A}\times\mathcal{B}$ when the inputs are $(x,y) \in \mathcal{X}\times\mathcal{Y}$. Since, by assumption, the inputs are uniformly distributed, $p^n(x,y) = 1/|\mathcal{X}\times\mathcal{Y}|$ for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$.

Now, for all $n \geq n_0$, the fact that A and B are deterministic boxes gives us,

$$p^n(a,b|x,y) = \delta_{A'(x,n)=a}\delta_{B'(y,n)=b},$$

a local deterministic distribution. Hence, for all $n \geq n_0$, by equation (3.1) we have that

$$E(B_N^n) = \frac{1}{|X\times\mathcal{Y}|}\sum_{a,b,x,y} B_{a,b,x,y}\delta_{A'(x,n)=a}\delta_{B'(y,n)=b} \leq \frac{B_L}{|X\times\mathcal{Y}|}.$$

Finally,

$$\begin{aligned} \lim_N E(B_N) &= \lim_N \frac{|\mathcal{X}\times\mathcal{Y}|}{N}\sum_{n=1}^{N} B_N^n \\ &= \lim_N \frac{|\mathcal{X}\times\mathcal{Y}|}{N}\left(\sum_{n=1}^{n_0} B_N^n\right) + \frac{|\mathcal{X}\times\mathcal{Y}|}{N}\left(\sum_{n=n_0}^{N} B_N^n\right) \\ &\leq B_L \end{aligned}$$

This concludes the proof. □

3.1.1 Relationship to standard hidden-variable models

Jarret [Jar84] showed that the factorizability condition (1.3) for local hidden-variable models (Definition 1.5.6) can be seen as the conjunction of two conditions: a condition called "Locality" by Jarrett and "Parameter Independence" (PI) by Shimony [Shi86], on the one hand, and a condition of "Completeness" (Jarrett) or "Outcome Independence" (OI) (Shimony), on the other.

Theorem 3.1.3 ([Jar84])**.** *A hidden-variable model $\langle\Lambda, p_\lambda, p\rangle$ is local iff*

$$p_\lambda(a|x,y) = p_\lambda(a|x) \text{ and } p_\lambda(b|x,y) = p_\lambda(b|y,) \tag{PI}$$

$$p_\lambda(a|b,x,y) = p_\lambda(a|x,y) \text{ and } p_\lambda(b|a,x,y) = p_\lambda(b|x,y) \tag{OI}$$

Corollary 3.1.4. *Every deterministic hidden-variable model for a non-local distribution* $\mathbf{p}$ *violates* (PI).

Proof. Every deterministic hidden-variable model $\langle \Lambda, q, A, B\rangle$ satisfies (OI). Indeed,

$$p(a|x,y,\lambda) = \delta_{a=A(x,y,\lambda)} \implies p(a|b,x,y,\lambda) = p(a|x,y,\lambda) \text{ and} \tag{3.2}$$
$$p(b|x,y,\lambda) = \delta_{b=B(x,y,\lambda)} \implies p(b|a,x,y,\lambda) = p(b|x,y,\lambda) \tag{3.3}$$

Hence, if $\sum_\lambda p(\lambda)p_\lambda(a,b|x,y)$ is non-local, it follows from Theorem 3.1.3 that it violates (PI). □

Corollary 3.1.4 can be interpreted as the distribution p_λ not satisfying the non-signaling constraints (Definition 1.5.10). Thus, we say that the model is signaling. As we stated before in the introduction, this does not imply the possibility of effectively signaling, as there are deterministic hidden-variable theories reproducing the quantum mechanical correlations [Boh52, TB03b]. In other words, the existence of signaling at the hidden-variable level does not imply the possibility of signaling at the observational level. The impossibility of effectively signal in this theories is a consequence of the model predictions being obtained from averaging over the hidden-variables. Furthermore, it was shown in [Val02] that every deterministic hidden-variable theory reproducing the quantum mechanical predictions for some distribution of the hidden-variables (labelled the "quantum equilibrium" distribution) predicts signaling correlations for other distributions (the "non-equilibrium" distributions).

Of course, the violation of (PI) in a non-local deterministic hidden-variable model should involve (at least) one hidden-variable occurring with non-vanishing probability.

Observation 3.1.5. *If* $\mathcal{M} = \langle \Lambda, q, A, B\rangle$ *is a deterministic hidden-variable model for a non-local distribution* $\mathbf{p}$*, then there exists* $\lambda \in \Lambda$ *with* $q(\lambda) > 0$ *such that*

$$\exists x\ \exists y, y'\ A(x,y,\lambda) \neq A(x,y',\lambda) \text{ or}$$
$$\exists x, x'\ \exists y\ B(x,y,\lambda) \neq B(x',y,\lambda).$$

Proof. By way of contradiction, suppose that for all $\lambda \in \Lambda$ such that $q(\lambda) > 0$ we have that

$$\forall x\ \forall y, y'\ A(x,y,\lambda) = A(x,y',\lambda) \text{ and}$$
$$\forall x, x'\ \forall y\ B(x,y,\lambda) = B(x',y,\lambda).$$

Then, the restriction of $\mathcal{M}$ to $\Lambda \setminus \{\lambda \in \Lambda \mid q(\lambda) = 0\}$ is a deterministic hidden-variable model for the non-local $\mathbf{p}$ satisfying (PI). A contradiction. □

Now, let us consider deterministic hidden-variable models in which the hidden-variable determining the outcomes at the n-th round of Bell experiment is, instead of sampled from some distribution q, a computable function of n.

Definition 3.1.6 (computable hidden-variable model)**.** A *computable hidden-variable model* $\langle \Lambda, \tilde{A}, \tilde{B}, f\rangle$ is a deterministic hidden-variable model $\langle \Lambda, q, A, B\rangle$ such that

$$q(\lambda) = \lim_n \frac{|\{i \leq n \mid f(i) = \lambda\}|}{n} \tag{3.4}$$

and $f : \mathbb{N} \to \Lambda$ (the pseudorandomness) is a computable function.

Proposition 3.1.7 below shows the relationship between this notion of computable hidden-variable models and the scenario we consider in this chapter.

Proposition 3.1.7. *If a non-local distribution* $\mathbf{p}$ *has a computable hidden-variable model, then there exists computable non-local boxes* $A : \mathcal{X} \times \mathcal{Y} \times \mathbb{N} \to \mathcal{A}$ *and* $B : \mathcal{X} \times \mathcal{Y} \times \mathbb{N} \to \mathcal{B}$ *such that*

$$p(a,b|x,y) = \lim_n \frac{|\{i \leq n \mid A(x,y,i) = a \wedge B(x,y,i) = b\}|}{n}$$

and

$$\exists^\infty n \; \exists \; x \; \exists \; y, y' \; A(x,y,n) \neq A(x,y',n) \text{ or} \tag{3.5}$$
$$\exists^\infty n \; \exists \; x, x' \; \exists \; y \; B(x,y,n) \neq B(x',y,n) \tag{3.6}$$

Proof. Let $\langle \Lambda, \tilde{A}, \tilde{B}, f \rangle$ be a computable hidden-variable model for a distribution $\mathbf{p}$. Define $A : \mathcal{X}\times\mathcal{Y}\times\mathbb{N} \to \mathcal{A}$ and $B : \mathcal{X}\times\mathcal{Y}\times\mathbb{N} \to \mathcal{B}$ as $A(x,y,n) \equiv \tilde{A}(x,y,f(n))$ and $B(x,y,n) \equiv \tilde{B}(x,y,f(n))$. Then,

$$\begin{aligned}
p(a,b|x,y) &= \sum_\lambda \left(\lim_n \frac{|\{i \leq n \mid f(i) = \lambda\}|}{n} \right) \delta_{A(x,y,\lambda)=a} \delta_{B(x,y,\lambda)=b} \\
&= \lim_n \sum_\lambda \frac{\sum_{i:f(i)=\lambda, i\leq n} \delta_{\tilde{A}(x,y,\lambda)=a} \delta_{\tilde{B}(x,y,\lambda)=b}}{n} \\
&= \lim_n \frac{\sum_{i\leq n} \delta_{\tilde{A}(x,y,f(i))=a} \delta_{\tilde{B}(x,y,f(i))=b}}{n} \\
&= \lim_n \frac{|\{i \leq n \mid \tilde{A}(x,y,f(i)) = a \wedge \tilde{B}(x,y,f(i)) = b\}|}{n} \\
&= \lim_n \frac{|\{i \leq n \mid A(x,y,i) = a \wedge B(x,y,i) = b\}|}{n}
\end{aligned}$$

Now, for all $\lambda \in \Lambda$ we have

$$p(\lambda) > 0 \text{ iff } \lim_n \frac{|\{i \leq n \mid f(i) = \lambda\}|}{n} > 0$$

and this implies

$$\exists^\infty n \; f(n) = \lambda.$$

Finally, combining this with by Observation 3.1.5 we have

$$\begin{aligned}
&\exists^\infty n \; \exists \; x \; \exists \; y, y' \; \tilde{A}(x,y,f(n)) \neq \tilde{A}(x,y',f(n)) \text{ or} \\
&\exists^\infty n \; \exists \; x, x' \; \exists \; y \; \tilde{B}(x,y,f(n)) \neq \tilde{B}(x',y,f(n))
\end{aligned}$$

and this concludes the proof. □

3.2 Using computable non-local boxes to signal

In the following, and for the sake of simplicity, we restrict to a 2-inputs-2-outputs Bell scenario, where $\mathcal{A} = \mathcal{B} = \mathcal{X} = \mathcal{Y} = \{0, 1\}$ (see Example 4). The extension to other scenarios is straightforward.

Let A and B be a pair of computable deterministic non-local boxes and, by Lemma 3.1.2, suppose, without loss of generality, that

$$\exists^{\infty} n \; \exists y \in \{0, 1\} \; B(0, y, n) \neq B(1, y, n), \tag{3.7}$$

This implies that, for infinitely many values of n, the value of x can be determined from the output of B with the suitable choice of y. Straightforwardly, if Alice knew how to compute B, they could trivially signal from Alice to Bob.

Proposition 3.2.1. *For every pair of computable non-local boxes A and B with B satisfying equation* (3.7)*, there is a protocol for Alice holding box A to send a message to Bob holding box B.*

Proof. Let $s \in \{0, 1\}^*$ be Alice's message, $eom \in \{0, 1\}^*$ an special end-of-message string and $u = s^\frown eom$. Let $[n_1, \ldots, n_{|u|}]$ be the first $|u|$ values of n such that $B(0, y_n, n) \neq B(1, y_n, n)$ for some y_n. Then, on round n_i, Alice inputs $u[n_i]$. Bob, on every round n such that $B(0, n, y_n) \neq B(1, n, y_n)$ for some y_n, will input such y_n and record $v[n] = 0$ if the output of his box was $B(0, n, y_n)$ and $v[n] = 1$ otherwise. When the last $|eom|$ bits of v equal eom, he stops. Alice's message will be in the first s bits of v. □

In the spirit of the device-independent formalism for Bell non-locality, the situation we want to study is when the players do not know the inner-workings of the boxes, i.e. they do not know how to compute A and B. In the following sections, using the tools of the theories of learnability of computable functions and computable randomness (Sections 1.7.1 and 1.8), we will show that:

Theorem (informal version). There exists a protocol such that, if Alice and Bob are holding computable non-local boxes,then they can perform one-way communication of any fixed size message provided they know a computable bound on the time computational complexity of the boxes.

The key idea of the protocol will be for Bob to perform a learnability in the limit scheme on the outputs of his box. Once function B is learnt, he could use the rounds n such that equation (3.7) holds to reconstruct Alice's input. There are three issues that we will need to deal with in this approach:

1. *Learning from incomplete samples.* In the traditional learning scenario, the learner, although only requiring a finite number of points $(x, f(x))$ to learn a target function f, he can, potentially, access the whole graph of f. In our non-locality scenario, however, the values $B(x_n, y_n, n)$ (i.e. the outputs of Bob's box) will, for every n, be known for just one pair $(x_n, y_n) \in \{0, 1\}^2$.

2. *Learning from distributed inputs.* Furthermore, Bob will not know the corresponding x_n (they are Alice's inputs).

3. *Signaling in the limit.* Assuming we solve the preceding issues and we device a learning in the limit strategy, since Bob will, in general, will not be able to *effectively* tell when he has learnt, we will need to modify the strategy used in the proof of Proposition 3.2.1 to reconstruct Alice's message from the output of Bob's box.

To cope with issues 1 and 3, Alice and Bob will alternate between *learning* rounds and *signaling* rounds. The former are rounds in which they know both parties inputs (they are pre-established) and are used by Bob to learn function B. The later are rounds that are used to send a message from Alice to Bob assuming B is already known. Choosing the pre-established inputs in a *sufficiently random* manner will allow them to cope with the third issue.

3.2.1 The signaling protocol

As a first step of the protocol, Alice and Bob fix a computable function $T : \mathbb{N} \to \mathbb{N}$ and assume B is computable in $O(T(n))$-time; the protocol will fail if this assumption is false. To perform the aforementioned alternation between learning and signaling rounds, they will share a sequence S whose symbols are either a pair of bits, or an integer between 1 and m, where m is the length of the message that Alice wants to communicate. As anticipated before, Bob will be using a enumeration-learner $L_T : \mathbb{N} \to \mathbb{N}$ for the class of functions computable in $O(T(n))$-time (see Definition 1.7.6). On the learning rounds, Alice and Bob will input their boxes with a prearranged input pair and Bob will use the output of his box to, through L_T, update his guess for a program that computes B. On the signaling rounds, Alice will input her message and Bob, acting according to his current guess for a program for B, will choose, whenever possible, the input y that allows him to tell Alice's input x.

The protocol has thus four parameters: a computable time function T, a sequence

$$S \in \{(0,0),(0,1),(1,0),(1,1),1,\ldots,m\}^\omega$$

(which is the one shared by Alice and Bob to perform the switching between the two kinds of rounds), a number m which represents the size of the message that Alice wants to send to Bob and the number N of iterations of the protocol which will be fixed in advance.

All in all, here is the signaling protocol $\mathcal{P}(T, S, m, N)$:

---

1. Bob initializes $v[1 \ldots m]$ to 0^m and $\widetilde{B}$ to a TM that computes $f(x, y, n) := 0$.
2. For each round $n \leq N$:
   (a) **Learning round:** if $S(n) = (x, y)$, Alice inputs x and Bob inputs y. Furthermore, Bob sets his current guess $\widetilde{B}$ of a Turing machine that computes B to $L_T([\langle x_{i_1}, y_{i_1}, i_1, B(x_{i_1}, y_{i_1}, i_1)\rangle, \ldots, \langle x, y, n, B(x, y, n)\rangle])$, with i_k being the past learning rounds.

(b) **Signaling round**: if $S(n) = i \in \{1, \dots, m\}$, Alice inputs the ith bit of her message and Bob uses his current guess $\widetilde{B}$ of a program that computes B to see if there is a y such that $\widetilde{B}(0, y, n) \neq \widetilde{B}(1, y, n)$. If there is such y, he inputs it and records the output of his box in $v[i]$. If not, he inputs 0 and disregard his box's output.

3. The output of the protocol is the string v held by Bob.

In the next section we prove that, for a suitable choice of S and for sufficiently large N, performing this protocol with non-local boxes satisfying equation (3.7), allows Alice's message to be communicated to Bob.

3.2.2 Soundness

In the following, we let $T : \mathbb{N} \to \mathbb{N}$ be some computable function and $A, B : \{0,1\} \times \{0,1\} \times \mathbb{N}$ be computable boxes, with B computable in $O(T(n))$-time and satisfying equation (3.7).

Let us start with a simple observation which, informally, states that the learning process always converges.

Observation 3.2.2. *There exists an N_0 such that for all $n, m \geq N_0$, if $\widetilde{B}_n$ and $\widetilde{B}_m$ are the learner's hypothesis at rounds n and m respectively, then $\widetilde{B}_n = \widetilde{B}_m$. That is, the learning converges.*

Proof. This follows from the fact that L_T is an enumeration-learner for the class of function computable in $O(T(n))$-time. Let $\mathbf{T}$ be the first Turing Machine computing B in the enumeration used by L_T. If L_T reaches $\mathbf{T}$ in some learning round n, then it will output $\mathbf{T}$ on every learning round $m \geq n$. If it doesn't reach $\mathbf{T}$, it is only because it has stabilized in another $\mathbf{T}'$ appearing before $\mathbf{T}$ in the enumeration. Either way, it converges. □

The following lemma establishes sufficient conditions for the signaling protocol to be successful.

Lemma 3.2.3. *For every message u by Alice, if S is such that properties (P_1) and (P_2) below hold, then $u = \mathcal{P}(T, S, |u|, N)$ for sufficiently large N.*

(P_1) There exists N_0 such that for all $n \geq N_0$, Bob's candidate program $\widetilde{B}$ at stage n is correct, that is $\widetilde{B}(x, y, n) = B(x, y, n)$ for all $x, y \in \{0,1\}$. In other words, B is learnt in the limit with finite anomalies (Definition 1.7.8).

(P_2) For the k-th bit of Alice's message and for infinitely many n, $S(n) = k \in \mathbb{N}$ and $B(0, y, n) \neq B(1, y, n)$ for some $y \in \{0,1\}$, i.e. the signaling mechanism happens for infinitely many rounds.

Proof. When (P_1) holds, there exists N_0 such that choosing Bob's input according to $\widetilde{B}$ in the signaling rounds is reliable. If (P_2) also holds, then this reliable use of $\widetilde{B}$ will let Bob reconstruct every bit of Alice's message infinitely often. □

Now, whether (P_1) and (P_2) hold or not will depend on the choice of shared switching sequence S. In the next section we consider the case of S being sampled uniformly at random and in the section following we prove our main result with S being a computable computably random sequence.

Soundness when alternating randomly

First, let us see that the protocol works when the switching between learning and signaling rounds is done randomly.

Proposition 3.2.4. *If $S(i) \in \{0,1\}^2 \cup \{1,\ldots,m\}$ are independent and uniformly distributed random variables, then properties (P_1) and (P_2) hold.*

Proof. To see that (P_1) holds we proceed by contraposition. Suppose that the learning procedure stabilizes in one of the finitely many TMs $\mathbf{M}$ appearing before one computing B in the enumeration, and whose outputs differ from those of B in infinitely many inputs (x, y, m). This would imply that for almost all rounds n in which S dictates learning, that n is not one of the infinitely many m for which $\mathbf{M}(x, y, m) \neq B(x, y, m)$ for some (x, y). It is easy to see that the probability of this happening when choosing the learning rounds n at random is 0.

To see that (P_2) holds it suffices to observe that amongst the infinitely many n where (3.7) is true, the probability that S picks finitely many of them to signal the k-th bit of the message is zero. □

However, letting the $S(i)$ be independent and uniformly distributed random variables would make our argument too weak, as it would mean that Alice and Bob have access to randomness, a non-computable resource, to test models of nature that are assumed to use only computable functions. So, the question is:

(Q_1) Can we find a computable S such that (P_1) and (P_2) hold?

Soundness when alternating pseudorandomly

It is easy to see that choosing a too simple sequence for S will not work. For example, if S is chosen such that, it indicates learning in the odd rounds and signaling in the even, the learning could converge to a program that coincides with B in almost all odd positions but, for the even positions, it outputs, say, the negation of B (this program, of course, also runs in $O(T(n))$-time). The following Lemmas 3.2.5 and 3.2.6 show that letting S be a $T(n)$-random (Definition 1.8.5) sequence does the trick.

Lemma 3.2.5. *If S is $T(n)$-random with $T = \Omega(n^2)$, then $\mathcal{P}(T, S, m, N)$ verifies (P_1) for sufficiently large N.*

Proof. By Observation 3.2.2, let N_0 be such that for all $n \geq N_0$, let learner hypothesis for the target function B is some program $\widetilde{B}$. This means that for all $n \geq N_0$ and all $x, y \in \{0,1\}$, if $S(n) = (x, y)$ then $\widetilde{B}(x, y, n) = B(x, y, n)$, i.e. at least in the learning

rounds, $\widetilde{B}$ coincides with B (making the learner not change its hypothesis). Assume by contradiction that for infinitely many n

$$\exists x, y \in \{0,1\}.\widetilde{B}(x,y,n) \neq B(x,y,n). \tag{3.8}$$

Now, letting $g : \mathbb{N} \to \{0,1\}$ be defined as $g(n) = 1$ iff equation (3.8) is true, Γ as $\{1,\ldots,m\}^2$ and noting that from the assumption of B computable in $O(T(n))$-time it follows that g is computable in $O(T(n))$-time, we have by Proposition 1.8.8 that S is not $T(n)$-random. A contradiction. □

Lemma 3.2.6. *If S is $T(n)$-random with $T = \Omega(n^2)$, then $\mathcal{P}(T,S,m,N)$ verifies (P_1) for sufficiently large N.*

Proof. By equation (3.7) we have that for infinitely many n

$$\exists y \in \{0,1\}.B(0,y,n) \neq B(1,y,n).$$

Let $g : \mathbb{N} \to \{0,1\}$ be defined as $g(n) = 1$ iff equation (3.7) is true, and assume by way of contradiction that there exists $k \in \{1,\ldots,m\}$ such that for almost all n we have that if $S(n) = k$ then $g(n) = 0$. Then, letting $\Gamma = \{0,1\}^2 \cup \{1,\ldots,m\} \setminus \{k\}$ and noting that from the assumption of B computable in $O(T(n))$-time it follows that g is computable in $O(T(n))$-time, we have by Proposition 1.8.8 that S is not $T(n)$-random. A contradiction. □

Finally, combining Lemmas 3.2.5, 3.2.6 and 3.2.3 together with the fact that from a program for a computable $T : \mathbb{N} \to \mathbb{N}$ one can compute a $T(n)$-random sequence (Theorem 1.8.7), we can formalize the main result of this chapter with a positive answer to question (Q_1) above.

Theorem 3.2.7. *Let $T : \mathbb{N} \to \mathbb{N}$ be such that $T = \Omega(n^2)$. Then, Alice and Bob can individually compute a $T(n)$-random sequence S such that for every message $u \in \{0,1\}^*$ by Alice, if they perform protocol $\mathcal{P}(T,S,|u|,N)$ using computable non-local boxes $A, B : \{0,1\} \times \{0,1\} \times \mathbb{N}$ with B computable in $O(T(n))$-time and satisfying 3.7, then $u = \mathcal{P}(T,S,|u|,N)$ for sufficiently large N.*

It is important to note that, without any knowledge of B, there is no a priori bound on the number of iterations N Alice and Bob will have to perform in order for her message to be communicated. Nonetheless, since this number is finite, there exists some finite distance for which the signaling allowed by our protocol is superluminal. For instance, if it takes M rounds for Bob to find out Alice's message and each round takes a time T, then if they are at a distance cTM, the message is obtained before a light signal from Alice could reach Bob. It could be argued that imposing a bound on the time complexity of Alice and Bob's boxes (which are nothing but an abstraction of what Nature is doing to choose the outputs) is a strong requirement. However, as we already mentioned in Chapter 2, since the number of computational steps per second that can be performed by a system of mass m is upper bounded by $2mc^2/\pi\hbar$ [Llo00], this is not only a requirement of our protocol but a reasonable physical assumption.

3.3 Discussion

Our protocol shows that correlated systems that would have violated a Bell inequality if were used for a standard Bell test (i.e., with random inputs), can be used to signal if assumed to be computable and a computable time bound for their computational complexity is known in advance. The main consequence of this is that we are left with the following consequences: either Bell-violating systems cannot be computable, or if Alice and Bob guess properly a complexity class larger than the one used by the computable systems, they can signal in either way using the previous protocol.

The only assumptions to arrive at this result were the computable nature of the boxes and the requirement of violating a Bell inequality if used for such matter.

It is worth mentioning that our model, in order to produce Bell inequality violating boxes, needs to use an internal signaling as a resource (Alice's box needs to know about Bob's input and viceversa). As we mentioned, this does not imply that Alice and Bob can send information to each other since this signaling doesn't necessarily reach the observational level. Also, if we are to analyse the computable nature of the outputs from simple quantum experiments (e.g. measuring some observable to a single qubit), and we are to extend such analysis to non-local boxes, there is no way out of this assumption.

This work shows that in device independent scenarios, computability of outputs imposes a strong limitation on how nature can behave if it only had computable resources to generate outputs for the experiments. Our result imply that, under the well established assumption that no observable signaling exists, we need to accept the existence of physical processes with uncomputable outputs.

It is worth mentioning that our result doesn't go into conflict with the different interpretations of quantum mechanics. All of them predict random outputs, which are not allowed by our model. In the Copenhagen interpretation, the measurement process is postulated as random, whereas, for example, in Bohmian mechanics, it is deterministic but the initial conditions are randomly distributed and fundamentally unknowable.

RESUMEN DEL CAPÍTULO

Es una consecuencia del teorema de Bell [Bel64] que cualquier explicación mediante una teoría de variables ocultas determinística de las correlaciones no-locales que la teoría cuántica predice y que hoy estamos prácticamente seguros que la Naturaleza exhibe [HBD$^+$15, GVW$^+$15, SMSC$^+$15], tiene que permitir la existencia the algún tipo de mecanismo de señalización que vincule elecciones y resultados de mediciones distantes. Pero, como las correlaciones cuánticas son no-señalizantes, tal mecanismo debe estar restringido al nivel de variables ocultas y no llegar al nivel fenomenológico.

Algunos ejemplos de explicaciones determinísticas de las correlaciones no-locales son: el modelo de variables ocultas con comunicación de Toner y Bacon [TB03b], y, de una manera más prominente, la mecánica de Bohm [Boh52]. Para aquellos modelos que usan comunicación clásica para simular no-localidad, uno de hecho puede estudiar la cantidad de comunicación necesaria (ver, por ejemplo, [RT09, SZ08, DKLR11]). En todas estas teorías, a pesar de que las salidas en cada ronda de un experimento de Bell están *determinadas* dadas las entradas y las variables ocultas, la variable oculta se elige al azar siguiendo alguna distribución de probablidad no deterministica.

En este capítulo estudiamos la clase de modelos determinísticos de las correlaciones no-locales en los cuales la elección de variable oculta es, en lugar de aleatoria, pseudoaleatoria. En principio, la secuencia de variables ocultas para un dado experimento es experimentalmente inaccesible; nos preguntamos si el hecho de ser computable tiene alguna consecuencia observacional.

Nuestro resultado principal es mostrar que todo modelo de variables ocultas determinístico de las correlaciones no-locales tiene que ser no-computable si queremos prevenir que tales correlaciones pueden ser usadas para señalizar. En otras palabras, mostramos que si el modelo determinístico es computable, el mecanismo de señalización escondido usado para exhibir no-localidad se puede extraer al nivel observacional y usado para comunicar información entre partes distantes siempre y cuando se conozca una cota superior a su complejidad temporal (**Theorem 3.2.7**). Más específicamente, damos un protocolo para realizar comunicación unidireccional entre dos observadores portando cajas no-locales en una clase de complejidad computacional conocida [BdlTS$^+$17].

Nuestro resultado implica que, en escenarios independientes-del-dispositivo, la computabilidad de las salidas impone una fuerte limitación a cómo puede comportarse la Naturaleza si sólo tiene recursos computables para generar la salida de los experimentos. Más precisamente, bajo la bien establecida hipótesis de que no existe señalización observable, uno tiene que aceptar la existencia de procesos físicos con salidas no-computables.

Es importante mencionar que nuestro resultado no entra en conflicto con las diferentes interpretaciones de la mecánica cuántica; todas ellas predicen salidas aleatorias,

algo no permitido por nuestro modelo. En la interpretación de Copenhague, el proceso de medición se postula aleatorio; mientras que, por ejemplo, en la mecánica de Bohm, es determinístico, pero las condiciones iniciales están distribuidas al azar y son fundamentalmente inconocibles.

4. PSEUDORANDOM MIXTURES OF QUANTUM STATES

With the advance of the experimental realization of quantum protocols, the most widely used kind of setups consist of classical systems controlling quantum ones [PWT$^+$07, TMF$^+$13, BCS$^+$04, TDH$^+$05]. Being classical, the control systems are limited in what they can achieve and this reflects on what can be achieved by the setups they control. In this chapter we consider this problem in the context of mixed state preparations. We will study different scenarios in which if pseudorandomness is used instead of randomness (as done in e.g. [AB09a, LKPR10]), situations which were initially indistinguishable become so. First, we show that a player (Bob) can distinguish, in finite time and with arbitrarily high probability, whether the qubits that another player (Alice) is preparing for him have been pseudorandomly chosen from the σ_z basis or from the σ_x basis; something impossible if she were picking them at random (**Theorem 4.1.7**). Notice that this, hence, implies that it is incorrect to characterize Bob's lack of knowledge about the preparation basis with the maximally mixed state [BdlTS$^+$16]. We provide the results of an experimental proof-of-concept of a special case of this result done by the group of Dr. Miguel Larotonda [LGSdlT$^+$]. Next, we generalize this result to any fixed initial preparation basis. Finally, we further extend the result to the situation in which, instead of having a fixed preparation basis, Alice is allowed to prepare any qubit state (with rational coefficients) (**Theorem 4.2.4**) [LGSdlT$^+$].

4.1 The basic scenario

The first scenario we will consider is described by the following game.

Definition 4.1.1 (basis-distinguishing game)**.** Let $1/2 > \delta > 0$ and $\mathcal{C} \subseteq \{0,1\}^\omega$ a class of computable sequences. The *basis-distinguishing game* is as follows. At the beginning, Alice picks a sequence $Y \in \mathcal{C}$ and then flips a fair coin to choose between σ_z and σ_x. Then, she uses the sequence Y to, upon Bob's n-th request, prepare to him the qubit state $|0\rangle$ (resp. $|+\rangle$) if $Y(n) = 0$ or the state $|1\rangle$ (resp. $|-\rangle$) if $Y(n) = 1$, when the initially chosen operator was σ_z (resp. σ_x). Bob's task is, by making quantum measurements on (finitely many of) Alice's qubits, to guess the preparation basis (i.e. either the eigenbasis of σ_x or the eigenbasis of σ_z) with a probability of error $P_{\text{err}} \leq \delta$.

In this chapter we will identify the values 1 and -1, obtainable when measuring the Pauli observables (Definition 1.4.6), to 0 and 1 respectively.

We will also consider a more experimentally inclined variation of this game in which the measurements made by Bob are noisy.

Definition 4.1.2 (noisy basis-distinguishing game)**.** Let $r \in [0,1]$. The *r-noisy basis-distinguishing game* is a basis-distinguishing game in which there is a probability r that when Bob measures in the preparation basis, the results are flipped.

We will study next how the game's difficulty depends on different choices of $\mathcal{C}$. But, before going into that, recall Observation 1.4.4 made in the Preliminaries:

Observation 1.4.4. *Two different ensembles can give rise to the same density matrix. For example, both the ensembles* $\{(1/2, |0\rangle), (1/2, |1\rangle)\}$ *and* $\{(1/2, |+\rangle), (1/2, |-\rangle)\}$ *give rise to the maximally mixed state* $\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

That is, if instead of using a computable sequence, Alice chooses the eigenstates by flipping a fair coin, no strategy allows Bob to distinguish the preparation basis.

Let us now consider the case in which $\mathcal{C}$ is a singleton, that is $\mathcal{C} = \{Y\}$ for some computable sequence Y. Of course, if Bob has access to (some) Turing machine computing Y, the game is trivial.

Proposition 4.1.3. *For every* δ *and every computable* $Y \in \{0,1\}^\omega$*, there is a strategy for Bob to win the finite basis-distinguishing game with probability of error* $P_{\text{err}} \leq \delta$ *when* $\mathcal{C} = \{Y\}$.

Proof. Given an error δ, Bob requests from Alice k qubits, with $k = \min_n[2^{-n} \leq \delta]$, which he then measure in the σ_z basis, generating a string z of length k with the measurement results. If the preparation basis is σ_z, then $z = Y \upharpoonright n$, and if the preparation basis is σ_x, then z is a k-bits string sampled uniformly at random. Then, if z and $Y \upharpoonright k$ coincide, he claims that the preparation basis is σ_z, otherwise he claims that it is σ_x. He makes an error when the preparation basis is σ_x and the string of measurement results coincide with $Y \upharpoonright k$. Therefore, $p_{err} = 2^{-k} \leq \delta$. □

Notice that the same kind of strategy applies if $\mathcal{C}$ is any finite class of computable sequences.

Corollary 4.1.4. *For every* δ *and every finite* $\mathcal{C}$*, there is a strategy for Bob to win the finite basis-distinguishing game with* $P_{\text{err}} \leq \delta$.

Proof. The strategy is essentially the same as the one given in the proof of Proposition 4.1.3, letting $k = \min_n[|\mathcal{C}|2^{-n} \leq \delta]$ and having Bob claim that the basis is σ_z if z matches the first k bits of any of the sequences in $\mathcal{C}$ and claim that it is σ_x otherwise. □

Next, let us consider the more interesting problem of trying to come up with a strategy that works when $\mathcal{C}$ is the class of all computable sequences. The above strategy will, of course, not work because Bob would have to compare prefix z with the first k bits of the infinitely many computable sequences (let alone that the class of all computable sequences is not computably enumerable). The first result of this section is:

Theorem 4.1.5. *For every* δ*, there is a strategy for Bob to win the finite basis-distinguishing game with probability of error* $P_{\text{err}} \leq \delta$ *when* $\mathcal{C}$ *is the class of all computable sequences.*

As the strategies outlined above, the one we will describe next to prove Theorem 4.1.7 can be divided in two parts: a (quantum) measurement part and a (classical) processing-of-the-measurement-outcomes part. The measurement part is as follows.

Measurement part

Bob measures σ_z to every qubit that Alice prepares on an even request number and σ_x to every qubit she preapares on an odd request number, yielding two binary sequences of measurement results Z and X respectively, as can be seen in Figure 4.1.

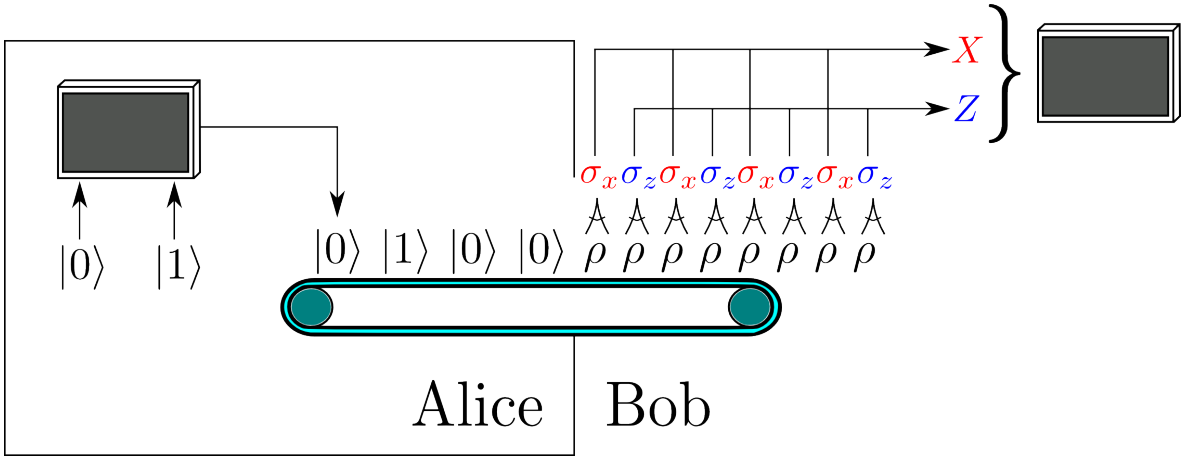

Fig. 4.1: Alice uses a computer to choose between $|0\rangle$ and $|1\rangle$ (or $|+\rangle$ and $|-\rangle$), keeping the basis fixed all through the experiment. To distinguish both possible preparations, Bob measures alternatively σ_x and σ_z and feeds the resulting sequences to a computer executing Algorithm 1.

The sequence corresponding to the choice of measurement that matches the preparation basis is computable (because it is either the odd or the even positions of the computable sequence Y Alice is using), and the other one, according to quantum mechanics, corresponds to a fair coin tossing, and so it is ML-random with probability 1. Therefore, we need an algorithm that given two sequences, one being computable and one arising from a fair coin tossing, is able to tell us which is which in finite time and with an arbitrarily high probability of success. For this, we will retort to the power of a universal ML-test (see Proposition 1.9.4). Before that, however, for ease of presentation, we will provide an explicit algorithm which, although not as general as the strategy using a universal ML-test, will be sufficient to win the game with arbitrarily high success probability while, at the same time, simplifying the explanation of the experiment in the next section. In the effective procedures we describe next the sequences X and Z must be understood as *oracles* [Soa99, §III].

Classical processing of the measurement outcomes: an explicit algorithm

To distinguish which of the two sequences X and Z is computable we *dovetail* between program number and maximum time steps that we simulate that program on the universal Turing machine $\mathbf{V}$ (that is, we simulate program 1 for 1 timestep, then programs 1 and 2 for 2 timesteps and so on), as is a common technique in computability theory. For each program p of length $|p|$ we will compare the first $k|p|$ output bits with the corresponding prefixes of both sequences, where $k \in \mathbb{N}$ will depend, as before, on the probability of success we are looking for. Whenever we find a match for the first $k|p|$

bits, we halt. The pseudocode for this effective procedure is in Algorithm 1.

Algorithm 1 The distinguishing protocol

Input: $k \in \mathbb{N}$ and $X, Z \in \{0,1\}^\omega$, one of them being computable
Output: 'X' or 'Z' as the candidate for being computable; wrong answer with probability bounded by $O(2^{-k})$

```
for t = 0, 1, 2 ... do
    for p = 0, ..., t do
        if V_t(p) = X ↾ k|p| then
            output 'X' and halt
        if V_t(p) = Z ↾ k|p| then
            output 'Z' and halt
```

Provided that at least one of X or Z is computable, the above procedure always halt —and so it only queries finitely many bits of both X and Z. Indeed, recalling that by Kleene's recursion theorem [Kle38] one can assume that TMs "know" their own index, we have that

Fact 4.1.6. For every $k \in \mathbb{N}$ and every computable sequence S there is a Turing machine $\mathbf{T}_e$ such that $\mathbf{T}_e(\epsilon) = S \upharpoonright k(e+1)$.

and therefore, in case $S \in \{X, Z\}$ is computable, for all $k \in \mathbb{N}$ there exists p such that $\mathbf{V}_t(p) = S \upharpoonright k \cdot |p|$ for some $t \in \mathbb{N}$.

Now, we bound the probability of having a miss-recognition, that is, the probability P_{err} that the above procedure outputs 'Z' when X was computable, or viceversa. To do so, we bound the probability that $S \in \{0,1\}^\omega$ has the property that for the given value of k there is p such that

$$(\exists t)\ \mathbf{V}_t(p) = S \upharpoonright k|p|. \tag{4.1}$$

Since there are 2^ℓ programs of length ℓ, the probability that there is a program p of length ℓ such that (4.1) holds is at most $2^\ell/2^{k\ell}$. Adding up over all possible lengths ℓ we obtain

$$P_{\text{err}} \leq \sum_{\ell>0} \frac{2^\ell}{2^{k\ell}} = \frac{2^{-(k-1)}}{1-2^{-(k-1)}} = O\left(2^{-k}\right), \tag{4.2}$$

which goes to zero with k going to infinity. Hence, by setting $k = \min_n[\frac{2^{-(n-1)}}{1-2^{-(n-1)}} \leq \epsilon]$, we have the desired bound on the probability of error P_{err}.

Noise robustness. Regarding the noisy version of the game (Definition 4.1.2), we will modify the algorithm so that it tolerates a fraction $q \in \mathbb{Q}$ of bit flips in the prefixes. The modified pseudocode is Algorithm 2, where d_H is the Hamming distance between two strings, which counts the number of different bits in both strings. The first thing to notice is that when $q = 0$ Algorithms 1 and 2 coincide.

We need to show now that, again, the success probability can be made as close to one as desired by choosing the parameter k. Instead of bounding the number of sequences that can be generated with a program of length ℓ, we need to bound the number of

Algorithm 2 The noise tolerant distinguishing protocol

Input: $q \in \mathbb{Q}$, $k \in \mathbb{N}$ and $X, Z \in \{0,1\}^\omega$, one of them being computable
Output: 'X' or 'Z' as the candidate for being computable; wrong answer with probability bounded by $O(2^{-k})$

```
for t = 0, 1, 2 ... do
    for p = 0, ..., t do
        if d_H(V_t(p), X ↾ k|p|) < qk|p| then
            output 'X' and halt
        if d_H(V_t(p), Z ↾ k|p|) < qk|p| then
            output 'Z' and halt
```

sequences that have a Hamming distance smaller than $qk\ell$ from a computable one. One possible bound is $2^\ell \binom{\ell k}{\lfloor q\ell k \rfloor} 2^{\lfloor q\ell k \rfloor}$, where the first exponential term counts the number of different programs of length ℓ, the combinatorial number corresponds to the number of bits that can be flipped due to errors, and the last exponential term gives which of these bits are actually being flipped. This estimation may not be tight, as we may be counting the same sequence several times. However, using this estimation we derive a good enough upper bound of the final error probability, as we get

$$P_{\text{err}} < \sum_{\ell>0} \frac{2^\ell 2^{\lfloor q\ell k \rfloor} \binom{\ell k}{\lfloor q\ell k \rfloor}}{2^{\ell k}}. \tag{4.3}$$

If we consider that $q < 1/2$, we can remove the integer part function and use the generalization of combinatorial numbers for real values. Then, by using that $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$, we obtain

$$P_{\text{err}} < \sum_{\ell>0} \left[2^{(1+qk-k)} \left(\frac{e}{q}\right)^{qk} \right]^\ell. \tag{4.4}$$

This geometric sum can be easily computed yielding

$$P_{\text{err}} < \frac{2^{1+qk-k} \left(\frac{e}{q}\right)^{qk}}{1 - 2^{1+qk-k} \left(\frac{e}{q}\right)^{qk}}. \tag{4.5}$$

Now, it can be shown numerically that for $q \lesssim 0.21$ the probability of mis-recognition tends to zero exponentially with k.

Finally, we show that (with probability 1) Algorithm 2 halts for all inputs satisfying the assumptions. Let $r < q$ be the probability of a bit flip. With probability 1, we have that for every δ there exist an m_0 such that for every $m > m_0$ the portion of bit flips in both $X \upharpoonright m$ and $Z \upharpoonright m$ are less than $(r + \delta)m$. This means that if we go to long enough prefixes (or programs), the portion of bit flips will be less than q. And since any computable sequence is computable by arbitrarily large programs, this ensures that our algorithm will, at some point, come to an end.

In the next section we give an alternative effective procedure for the classical processing stage of the distinguishing protocol using ML-tests which is shorter and more general, albeit possibly more technically involved. In Section 4.3 we provide a proof-of-concept implementation of a (simplified version) of Algorithm 2.

Classical processing of the measurement outcomes: using a universal ML-test

Let $(U_m)_{m\in\mathbb{N}}$ be a universal ML-test (Proposition 1.9.4) and let $k = \min_m[2^{-m} \leq \epsilon]$. Bob starts enumerating all the strings in $U_k = \{\sigma_1, \sigma_2, \dots\}$ until he finds some n such that for $Y = X$ or $Y = Z$ we have

$$[Y \upharpoonright n] \subseteq \bigcup_{i\leq n}[\sigma_i].$$

Since either X or Z is computable, the last condition has to be satisfied for sufficiently large n. If the above condition was first satisfied by $Y = X$, he claims that X is the computable sequence and that Y is the random one; if the above condition was first satisfied by $Y = Z$ he claims he claims that Z is the computable sequence and that X is the random one. This decision is wrong when the random sequence was captured by $[U_k]$ before the computable one was (of course, for some $k' > k$ the random sequence would be out of $[U_{k'}]$). Hence, the probability of making this error is at most the probability for the coin tossing sequence to be inside $[U_k]$, and this is, by definition, at most 2^{-k}.

Observe that in the above protocol there is nothing special with one of the sequences being computable. All that matters is that one of the sequences is not ML-random (of which the computable sequences are, of course, a subset).

Noise robustness. As we did in the preceding section, let us consider the noisy version of the distinguishing game. Having a flip probability of r means that the sequence of measurement results when measuring in the preparation basis is Yxor N, where the xor is taken bitwise and N, the noise sequence, is an infinite sequence such that , with probability 1, the limit relative frequency of the symbol 0 is strictly greater than the expected value, i.e.

$$\limsup_n \frac{\#\{i \leq n \mid N(i) = 1\}}{n} = r < 1/2.$$

Therefore, N is not ML-random since it does not satisfies the law of large numbers (see Proposition 1.8.6), and if Y is computable then Yxor N is also not ML-random. Now, Bob can apply the same protocol as above to distinguish Yxor N, which is not ML random, from the one coming from the coin tossing. Thus, we have shown that:

Theorem 4.1.7. *For every r and δ there is a strategy for Bob to win the r-noisy basis-distinguishing game with probability of error $P_{\text{err}} \leq \delta$ when $\mathcal{C}$ is the class of all total computable functions.*

4.1.1 Distinguishing any (initially fixed) preparation basis

Now, consider the following slightly more general scenario. Player Bob is presented with two boxes, B_1 and B_2, with the promise that one of them prepares qubits in the maximally mixed state and the other prepares states pseudorandomly chosen from a fixed basis $\mathbb{B}$ known to Bob. His task is to distinguish which box is which in finite time and with arbitrarily low probability of error. It is easy to see that a slight modification of the winning strategy for the finite basis-distinguishing game allows him to also win in this scenario. Namely, if instead of alternating between measuring σ_x and measuring σ_z as in Section 4.1, Bob measures the outputs of both boxes in the $\mathbb{B}$ basis, the binary sequence sequence associated with the box which has the computer will be computable and the other, according to quantum mechanics, independents tosses of a fair coin and so Martin-Löf random. Hence, with any of the classical post-processing protocols outlined above, he is able to distinguish both situations. In the next section we further generalize the result to the scenario in which the box using pseudorandomness is not restricted to choosing states from a fixed qubit basis.

4.2 Distinguishing any pseudorandom mixture of qubits

The fully general scnenario is described by the following game.

Definition 4.2.1 (find-the-identity game)**.** Let $1/2 > \delta > 0$. The *find-the-identity game* is as follows. Player Bob is presented with two boxes: B_1 and B_2. One of the boxes prepares single qubit maximally mixed states (a.k.a. the identity); the other box contains a computer producing, at each round n, rational numbers $\theta_n, \phi_n \in [0, 2\pi]$ and preparing a qubit in the state $|\psi_n\rangle = \cos(\theta_n/2)|0\rangle + e^{i\phi_n}\sin(\theta_n/2)|1\rangle$. Bob's task is, by making quantum measurements on (finitely many of) the qubits coming out of the boxes, to distinguish which box is which with a probability of error $P_{\text{err}} \leq \delta$.

The main result of this section is a protocol for Bob to win this game with arbitrarily high probability and independently of the program being run by the computer.

Bob's protocol works as follows. In each round he will randomly pick between σ_x, σ_y and σ_z, using for instance a QRNG, and measure such observable to the qubit coming out from each box. This gives rise to three sequences:

1. A ternary sequence M formed by the measurement choices performed in each round. Formally, $M(i) = 1$ if Bob measures σ_x at the i-th round (resp. 2 for σ_y and 3 for σ_z).

2. A binary sequence B_1 formed by the results of the measurements over the qubits coming from box number 1. Formally, $B_1(i) = 0$ (resp. $B_1(i) = 1$) if the result of measuring the observable represented by $M(i)$ to the qubit coming out of box 1 at round i was 0 (resp. 1).

3. Analogously, a binary sequence B_2 formed by the results of the measurements over the qubits coming from box number 2.

Note that, as in the basis-distinguishing game, although Bob measures finitely many times, the sequences are potentially infinite in the sense that he can keep requesting qubits from both boxes and making as many measurements as he needs.

As we will see now, sequences B_1 and B_2 have a distinctive feature that will allow Bob to distinguish which is the maximally mixed state and which is the one being produced by a computer.

Let $r \in \{1, 2\}$ be the box preparing the maximally mixed state and $c = 3 - r$ be the box with the computer inside. With probability 1, the sequence B_r will be Martin-Löf random with respect to sequence M. This follows from the fact that, irrespective of the measurement basis $M(i)$, $B_r(i)$ is a fair coin tossing for all i. On the other hand, sequence B_c will not be Martin-Löf random with respect to M. This is not straightforward, and we prove it next. First we need the following Lemma:

Lemma 4.2.2. *Let $|\psi_j\rangle$ and σ_j be, respectively, the pure state produced by box c and the observable corresponding to Bob's choice of measurement, for round j. With probability at least $1/3$ we have $|\langle\psi_j|\sigma_j|\psi_j\rangle| > 0.1$.*

Proof. This follows from the fact that Bob chooses from $\{\sigma_x, \sigma_y, \sigma_z\}$ uniformly at random and that every pure state gives a biased result in either of the three measurement choices (that is, for all $|\psi\rangle$ at least one of $|\langle\psi|\sigma_x|\psi\rangle| > 0.1$, $|\langle\psi|\sigma_y|\psi\rangle| > 0.1$ or $|\langle\psi|\sigma_z|\psi\rangle| > 0.1$ holds). □

This will allow us to prove a second Lemma which will let us conclude that any computably preparation made by Alice is distinguishable from the correctly prepared maximally mixed state.

Lemma 4.2.3. *With probability 1, sequence B_c is not ML-random relative to M.*

Proof. Following Lemma 4.2.2, let us assume, without loss of generality, that for infinitely many rounds n, the probability of obtaining 1 when measuring the state prepared by Alice in the direction $M(n)$ is greater than 0.55. This means that there is an effective way, using M as an oracle, to identify a subsequence Y of B_c not satisfying the law of large numbers (with probability 1). Namely, let $h : \mathbb{N} \to \mathbb{N}$ be defined as

$$h(n) := \min_m \left[[\mathrm{tr}(\Pi_1^{(m)}|\psi_m\rangle\langle\psi_m|) > 0.55] \wedge [\forall i < n \; m > h(i)] \right]$$

with $\Pi_1^{(n)}$ the projector to the eigenspace of eigenvalue 1 of observable $M(n)$. We have, with probability 1, that

$$Y = B_c(h(0))B_c(h(1))B_c(h(2))\ldots$$

does not satisfy the law of large numbers and therefore, by Proposition 1.8.6, is not ML-random. Hence, noting that $|\psi_m\rangle$ is computable from m (e.g. with Alice's program) and so h is computable relative to M, we have that, by relativizing Proposition 1.8.4, B_c is not ML-random relative to M with probability 1. □

We have proven so far that, with probability 1, B_c is not ML-random relative to M but B_r is. This fact, together with the existence of an universal oracle Martin-Löf test $(U_m^M)_{m\in\mathbb{N}}$, implies that Bob has an effective procedure using his sequence of

measurement choices M as an oracle to distinguish, with arbitrarily small probability of error, which of the boxes is using a computer to prepare its states. Namely, given a significance level 2^{-m}, he starts enumerating all the strings in $U_m^M = \{\sigma_1, \sigma_2, \dots\}$ until he finds some n such that

$$[B_i \upharpoonright n] \subseteq \bigcup_{i \leq n} [\sigma_i].$$

for some $i \in \{1, 2\}$ and claims that box i is the one with the computer. Since, with probability 1, either B_1 or B_2 is not ML-random relative to M, the last condition has to be satisfied for sufficiently large n with probability 1. This decision is wrong when the sequence ML-random relative to M was captured by $[U_m^M]$ before the non-ML-random one was (of course, for some $m' > m$ the random sequence would be out of $[U_{m'}^M]$). Hence, the probability of making this error is at most the probability for the coin flipping sequence to be inside $[U_m^M]$, and this is at most 2^{-m}. Therefore, we have shown that:

Theorem 4.2.4. *For every $\delta > 0$ there is a randomized strategy for Bob to win the find-the-identity game with probability of error $P_{\text{err}} \leq \delta$.*

4.3 Proof-of-concept experiment for the basis-distinguishing game

In this section we reproduce the results of a proof-of-concept experiment done by Miguel Larotonda and Ignacio López Grande of the winning strategy for the noisy basis-distinguishing game which uses Algorithm 2.

Algorithm 2 searches the whole space of all Turing machines and thus it is, of course, impossible in practical terms. Therefore, for this experiment, we have restricted the set of computable sequences used by Alice to those which are the output of the *rand* function from Matlab, using the Mersenne twister default generator algorithm [MN98] and with a initial seed of fixed maximum length. Of course, this means that we are back in the situation of Corollary 4.1.4-finite-set-of-sequences and therefore, no alternation between measurement basis is required. However, to have the proof-of-concept be as faithful as possible, we will nevertheless perform the alternation as outlined above. Finally, some minor changes to Algorithm 2 were required due to the non-deterministic nature of the emission and detection of *Poissonian* single photon states used as physical implementation for qubits. The adapted protocol can be specifically stated as follows:

- Alice and Bob set the value of two parameters from the protocol: ℓ_{max} which determines the maximum length of the *rand* function seed to be used and k which bounds to N = $k \times \ell_{max}$, the number of qubits to be transmitted on any run of the experiment.
- Alice pseudorandomly chooses one integer between 0 and $2^{\ell_{max}}$-1 which is used as the initial value, or seed for the *rand* function. The output of *rand* is binarized using the *round*() function resulting on a string of N *pseudorandom* bits.
- Alice chooses randomly (with fair coin randomness as explained below) the basis in which she will encode and send the string.
- Alice sends the N qubits to Bob. She encodes the binary string information in the photon polarization degree of freedom of a faint pulsed light beam.

- Bob measures the $\frac{N}{2}$ even and $\frac{N}{2}$ odd elements, each in one of the mutual unbiased bases of σ_x or σ_z.
- Bob, after measurement, computes the Hamming distance (for even and odd bits) between experimental data and the output of *rand*() function with the different seeds. When the minimum Hamming distance condition is fulfilled Bob ends the search.
- Finally Bob compares the state preparation (σ_x or σ_z mixtures) predicted by him with the mixture that was actually prepared by Alice to estimate the error probability (P_{err}) of the prediction.

A complete experiment consists in several repetitions of the protocol sketched above. Every execution is divided in two parts; the *transmission* of qubits from Alice to Bob, followed by a *search* routine, where Bob compares both bit strings with the strings generated by the *rand* function over all the possible seeds of length bounded by ℓ_{max} as it is stated in the theoretical protocol. When Bob finds a string that resembles the experimental series up to a certain Hamming distance value, the search ends. The result is compared with the actual basis used by Alice and the wrong guesses are registered as errors. After this they repeat the procedure with a new seed pseudorandomly picked, and a new random emission basis choice. The bound for the Hamming distance allows us to control the tolerance of the experiment against the Quantum Bit Error Rate (QBER).

One thing to be noticed is that Bob may not find a series that fulfills the desired Hamming distance condition. This is a situation that is not present in the theoretical protocol. In this way every time that Bob doesn't find a match we compute the experiment as inconclusive and it is discarded. To overcome this issue, the parameters of the protocol (such as maximum Hamming distance allowed) were set to guarantee that the probability of error occurrence was always greater than the probability of not finding any bit string fulfilling the condition. Under such assumptions, and using reasonable tolerances, we find that the ratio of inconclusive experiments to total number of errors was negligible.

The experiment involved 3100 repetitions of the *transmission* and *search* protocols. The total number of qubits transmitted on each repetition was fixed, and set by $k_{max} \times \ell_{max}$ (in this implementation $\ell_{max} = 10$). The parameter k determines the theoretical error probability for a given tolerance (q) and was set to take values between 1 and 16. This bounds the maximum number of compared bits on each Hamming distance calculation to $N = 320$ ($\ell_{max} \times k_{max}$ bits for even and odd bits); that is the number of qubits that Alice sends to Bob on each run.

After the qubit transmission is finished, Bob begins the search procedure building a list of $\sum_{i=1}^{\ell_{max}} 2^i = 2^{\ell_{max}+1} - 2$ "programs" (i.e. binary seeds to the *rand* function). Of course, since the seed to the *rand* function is ultimately an integer, different "seed strings" in the list (e.g. 0 and 00) will produce the same output of a run of the function; what will be different is the length of subsequence of the binarized output we will use to compare with the measurement data (in the e.g. before, k and $2k$). When either the even or odd bits of the compared strings fulfil the Hamming Distance criterion. Finally he compares the basis for the mixed state preparation predicted by this protocol with the one that Alice actually used, for the error probability estimation.

4.3.1 Experimental setup

The above protocol was put to the test on a photonic setup, based on a modified BB84 Quantum Key Distribution (QKD) implementation [LGSL16] which consists of an emission

stage that is able to send binary states coded in two different unbiased bases of the photon polarization, which are called computational basis and diagonal basis, and a reception stage for the quantum channel. Additionally, a classical communication channel is added for synchronization, transmission and data validation. See Figure 4.2 for an schematic description of the setup.

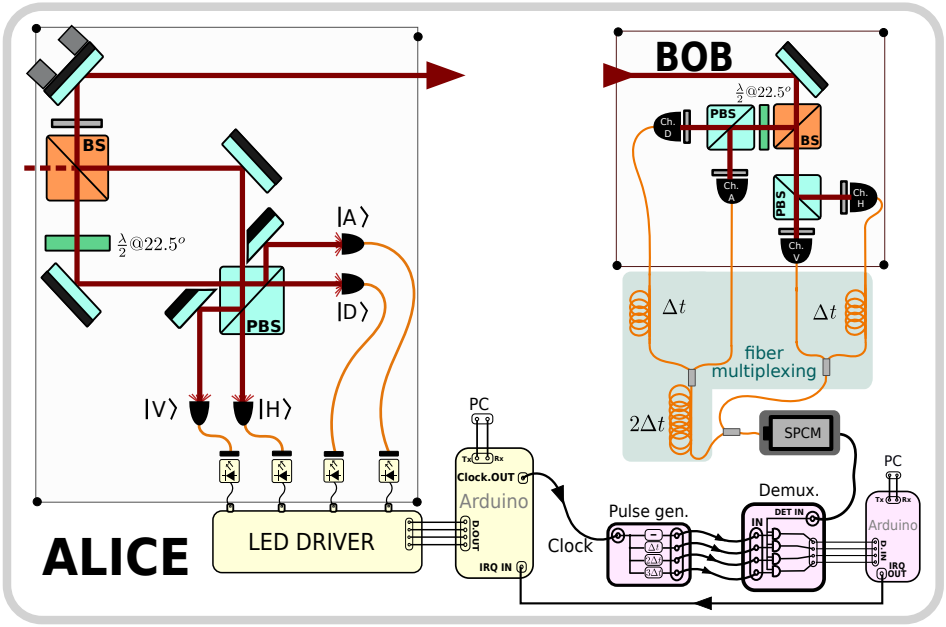

Fig. 4.2: Complete setup for implementing the transmission and search protocol: Qubits encoded in polarized faint pulses are produced by infrared LEDs. Light is coupled into and de-coupled from multimode fibers to obtain uniform beams for the four sources. The polarization state preparation is achieved by passing through a PBS (for H and V states) and an extra halfwave-plate for the D and A paths. A non-polarizing Beam Splitter cube couple the optical paths into an only exit light path. At the receiver's side a BS passively and randomly selects the detection basis for each incoming pulse. The outputs are coupled into multimode optical fibers, where different delays are imposed to make a polarization to time-bin transformation into a common output fiber. Finally a photon counter module and a temporal mask demultiplexer are used for detection.

4.3.2 Complete Results and Simulations

Herein we analyze the experimental results. We compare the performance of Bob at guessing the emission basis, with the theoretical error probability P_{err} (Equation (4.5)), and we also present additional data analysis aiming to explain the behavior of the error rate obtained.

As a result of each run, Bob gets two 160-bit length strings. M_e are the outcomes of even qubits, measured in the computational basis and M_o are the outcomes of odd qubits, measured in the diagonal basis. Bob compares these strings with the pair of strings from the program list S_e^j and S_o^j, where j stands for the number of program evaluated.

Note that when evaluating a program of length ℓ just the first $k \times \ell$ bits of the transmitted string are taken into account to compute the Hamming distance. The whole 160 bit string is only used in the Hamming distance measure of programs with $\ell_{max} = 10$.

The Hamming distance between the strings is calculated $H(S_e^j, M_e)$ and $H(S_o^j, M_o)$, and

the search finishes when one of them fulfills the tolerance criteria: $H(S_i^n, M_i) \leq \lfloor q \times k \times \ell(n) \rfloor$ from the noise tolerant protocol. In this experiment the tolerance parameter is set to $q = 0.15$. The result of the search for each run is registered for a further estimation of the error rate P_{err}.

The probability of error in Bob's guess of the emission basis can be estimated for different values of the parameter k. Figure 4.3 shows the error rate obtained from the experimental data and from a computational simulation of the experiment, together with the theoretical bounds for the distinguishing – noiseless and noise tolerant – protocols.

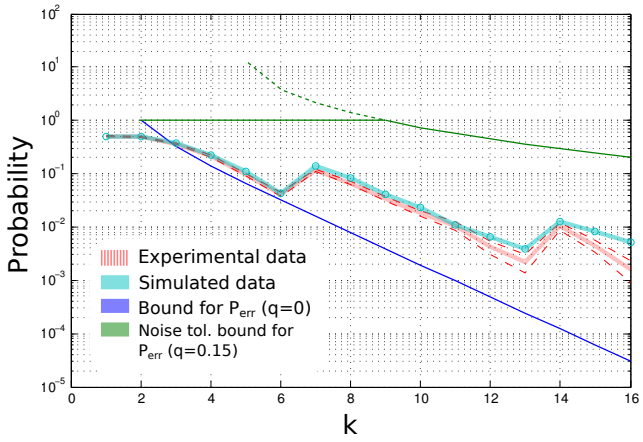

Fig. 4.3: The plot shows the experimental error rate obtained with the noise tolerant protocol (red lines), compared with the theoretical bounds for: the noiseless (blue line) and noise tolerant (green line) algorithms. The cyan line is the computational simulation of the experimental data taking into account the average QBER.

4.4 Discussion

We have shown that if Alice uses a computer to prepare a seemingly proper mixture of qubit states, Bob can distinguish it from the maximally mixed state in finite time, with arbitrarily high probability and without any access to Alice's algorithm. Additionaly, we have presented a proof-of-concept experiment showing that mixing two different sets of pure states that are supposed to yield the same mixed state, can be distinguished when mixed using the default pseudorandom number generator from Matlab.

Our distinguishing protocols, although impractical, fulfil their purpose of showing the distinguishability of situations which wouldn't be so if randomness were used instead of pseudorandomness. Our results imply that it is incorrect to model Bob's lack of knowledge in this scenario with independent copies of the maximally mixed state and they apply to, for instance, the mixed states experimentally produced using a classical random number generator [AB09a, LKPR10].

RESUMEN DEL CAPÍTULO

Con el avance de la realización experimental de protocolos cuánticos, las configuraciones más ampliamente utilizadas consisten de sistemas clásicos controlando sistemas cuánticos[PWT$^+$07, TMF$^+$13, BCS$^+$04, TDH$^+$05]. Siendo clásicos, los sistemas de control están limitados en lo que pueden conseguir, y esto se refleja en lo que puede ser logrado por las configuraciones que controlan. En este capítulo consideramos este problema en el contexto de preparación de estados mixtos.

En este capitulo estudiamos diferentes escenarios en los cuales, si se usa pseudoaleatoriedad en vez de aleatoriedad (como fue hecho, por ejemplo, en [AB09a, LKPR10]), situaciones que inicialmente eran indistinguibles se vuelven distinguibles. Primero, mostramos que un jugador (Bob) puede distinguir, en tiempo finito y con una probabilidad arbitrariamente alta, si los qubits que otro jugador (Alice) está preparando para él han sido pseudoaleatoriamente elegidos a patir de la base σ_z o la σ_x; algo que sería imposible si ella estuviera eligiéndolos al azar (**Theorem 4.1.7**). Observar que esto, por lo tanto, implica que es incorrecto caracterizar la falta de conocimiento de Bob sobre la base de preparación con el estado máximamente mixto [BdlTS$^+$16]. Proveemos los resultados de una prueba-de-concepto experimental de un caso especial de este resultado hecho por el grupo del Dr. Miguel Larotonda [LGSdlT$^+$]. A continuación, generalizamos este resultado a cualquier base de preparación inicial fija. Finalmente, extendemos adicionalmente el resultado a la situación en la cual, en vez de tener una base de preparación fija, Alice tiene permitido preparar cualquier estado de qubit (con coeficientes racionales) (**Theorem 4.2.4**) [LGSdlT$^+$].

5. ROBUST BELL VIOLATIONS FROM COMMUNICATION COMPLEXITY LOWER BOUNDS

The question of achieving large Bell violations has been studied since Bell's seminal paper in 1964 [Bel64]. In one line of investigation, proposals have been made to exhibit families of distributions which admit unbounded violations [Mer90, LPŻB04, NLP06, PGWP$^+$08]. In another, various measures of nonlocality have been studied, such as the amount of communication necessary and sufficient to simulate quantum distributions classically [Mau92, BCT99, Ste00, TB03a, Pir03, DKLR11], or the resistance to detection inefficiencies and noise. More recently, focus has turned to giving upper and lower bounds on violations achievable, in terms of various parameters: number of players, number of inputs, number of outputs, dimension of the quantum state, and amount of entanglement [DKLR11, JPPG$^+$10b, JP11].

Up until quite recently, violations were studied in the case of specific distributions (measuring Bell states), or families of distributions. Buhrman *et al.* [BRSdW12] gave a construction that could be applied to several problems which had efficient quantum communication protocols (Definition 1.10.7) and for which one could show a trade-off between communication and error in the classical setting. This still required an *ad hoc* analysis of communication problems. Recently Buhrman *et al.* [BCG$^+$16] proposed the first general construction of quantum states along with Bell inequalities from any communication problem. The quantum states violate the Bell inequalities when there is a sufficiently large gap between quantum and classical communication complexity (a super-quadratic gap is necessary, unless a quantum protocol without local memory exists).

We revisit the question of achieving large Bell violations by exploiting known connections with communication complexity. Strong lower bounds in communication complexity, equivalent to the partition bound [JK10], amount to finding *inefficiency-resistant Bell inequalities* [LLR12]. These are Bell functionals that are bounded above by 1 on all local distributions *that can abort*.

First, we study the resistance of normalized Bell inequalities to inefficiency. We show that, up to a constant factor in the value of the violation, any normalized Bell inequality can be made resistant to inefficiency while maintaining the normalization property (**Theorem 5.2.1**).

Second, we show how to derive large Bell violations from any communication problem for which the partition bound is bounded below and the quantum communication complexity is bounded above. The problems studied in communication complexity are far beyond the quantum set, but we show how to easily derive a quantum distribution from a quantum protocol. The Bell value we obtain is 2^{c-2q}, where c is the partition lower bound on the classical communication complexity of the problem considered, and q is an upper bound on its quantum communication complexity (**Theorem 5.3.2 and Corollary 5.3.3**). The quantum distribution has one extra output per player compared to the original distribution and uses the same amount of entanglement as the quantum protocol plus as many EPR pairs as needed to teleport the quantum communication in the protocol. We show that these Bell

violations can be made noise-resistant, at the cost of a 2^{2q} factor in the number of outcomes per player (**Theorem 5.4.1**).

Finally, we provide tools to build Bell inequalities from communication lower bounds in the literature. Lower bounds used in practice to separate classical from quantum communication complexity are usually achieved using corruption bounds and its variants. In **Theorem 5.5.3**, we give an explicit construction which translates these bounds into a suitable Bell functional. Table 5.3 summarizes the new results or the improvements that we obtain in this work.

Problem	Normalized Bell violations [BCG+16]	Inefficiency-resistant Bell violations (this work)
VSP [Raz99a, KR11]	$\Omega\left(\sqrt[6]{n}/\sqrt{\log n}\right)$ $d = 2^{\Theta(n \log n)}, K = 2^{\Theta(n)}$	$2^{\Omega(\sqrt[3]{n})-O(\log n)}$ $d = 2^{O(\log n)}, K = 3$
DISJ [Raz92, Raz03, AA05]	N/A	$2^{\Omega(n)-O(\sqrt{n})}$ $d = 2^{O(\sqrt{n})}, K = 3$
TRIBES [JKS03, BCW98]	N/A	$2^{\Omega(n)-O(\sqrt{n}\log^2 n)}$ $d = 2^{O(\sqrt{n}\log^2 n)}, K = 3$
ORT [She12, BCW98]	N/A	$2^{\Omega(n)-O(\sqrt{n}\log n)}$ $d = 2^{O(\sqrt{n}\log n)}, K = 3$

Tab. 5.1: Comparison of the Bell violations obtained by the general construction of Buhrman *et al.* [BCG+16] for normalized Bell violations (second column) and this work, for inefficiency-resistant Bell violations (see Propositions 5.5.4, 5.5.8, 5.5.11, and 5.5.14), in terms of the dimension d of the local Hilbert space, the size n of the of measurement settings (or inputs) sets (typically $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$) and the number of outcomes K (or outputs) per party of the quantum distributions. Explicit Bell inequalities are given in Section 5.5.2. The construction of Buhrman *et al.* only yields a violation when the gap between classical and quantum complexities is more than quadratic. In the case where the gap is too small to prove a violation, we indicate this with "N/A".

5.1 Background

5.1.1 Distributions that can abort

In this chapter, we will augment the output sets $\mathcal{A}$ and $\mathcal{B}$ with a special symbol $\perp \notin \mathcal{A} \cup \mathcal{B}$ which we call: the *abort* outcome. We will denote this with a superscript in the notation; namely $\mathcal{L}^{\perp}$, $\mathcal{Q}^{\perp}$ and $\mathcal{NS}^{\perp}$ will denote, respectively, the local, quantum and non-signaling distributions over outcomes in $\mathcal{A} \cup \{\perp\} \times \mathcal{B} \cup \{\perp\}$. The motivation for introducing outcome $\perp$ comes from the scenario considered in the detection loophole (see Definition 1.5.11) where we may have rounds in a Bell experiment where no measurement result is recorded. In a communication protocol, if the players output $\perp$ we say that they (or, equivalently, the protocol) aborts.

5.1.2 Measures of nonlocality

We have described nonlocality as a yes/no property, but some distributions are somehow more nonlocal than others. To have a robust measure of nonlocality, it should verify some common sense properties: for a fixed distribution, the measure should be bounded; it should also be convex, since sampling from the convex combination of two distributions can be done by first picking randomly one of the two distributions using shared randomness, and then sampling from that distribution. We also expect such a measure of nonlocality to have various equivalent formulations. Several measures have been proposed and studied: resistance to noise [KGZ$^+$00, ADGL02, PGWP$^+$08, JPPG$^+$10a], resistance to inefficiency [Mas02, MPRG02, LLR12], amount of communication necessary to reproduce them [Mau92, BCT99, Ste00, TB03a, Pir03, DKLR11], information-theoretic measures [BCSS11, GWAN12, FWW09], etc.

In the form studied in this chapter, normalized Bell inequalities were first studied in [DKLR11], where they appeared as the dual of the linear program for a well-studied lower bound on communication complexity, known as the nuclear norm ν [LS09] (the definition is given in Section 5.1.3). There are many equivalent formulations of this bound. For distributions arising from boolean functions, it has the mathematical properties of a norm, and it is related to winning probabilities of XOR games. It can also be viewed as a gauge, that is, a quantity measuring by how much the local set must be expanded in order to contain the distribution considered. For more general non-signaling distributions, besides having a geometrical interpretation in terms of affine combinations of local distributions, it has also been shown to be equivalent to the amount of local noise that can be tolerated before the distribution becomes local [JPPG$^+$10b].

A subsequent paper [LLR12] studied equivalent formulations of the partition bound, one of the strongest lower bounds in communication complexity [JK10]. This bound also has several formulations: the primal formulation can be viewed as resistance to detector inefficiency, and the dual formulation is given in terms of inefficiency-resistant Bell inequality violations.

In this chapter, we show how to deduce large violations on quantum distributions from large violations on non-signaling distributions, provided there are efficient quantum communication protocols for the latter.

5.1.3 Communication complexity lower bounds

To give upper bounds on communication complexity it suffices to give a protocol and analyze its complexity. Proving lower bounds is often a more difficult task, and many techniques have been developed to achieve this. The methods we describe here are complexity measures which can be applied to any function. To prove a lower bound on communication, it suffices to give a lower bound on one of these complexity measures, which are bounded above by communication complexity for any function. We describe here most of the complexity measures relevant to this work.

The nuclear norm ν, given here in its dual formulation and extended to non-signaling distributions, is expressed by the following linear program [LS09, DKLR11]. (There is a quantum analogue, γ_2, which is not needed in this work. We refer the interested reader to the definition for distributions in [DKLR11]).

Definition 5.1.1 ([LS09, DKLR11])**.** The *nuclear norm* ν of a non-signaling distribution

$\mathbf{p} \in \mathcal{C}$ is given by

$$\begin{aligned} \nu(\mathbf{p}) := \max_{B} \quad & B(\mathbf{p}) \\ \text{subject to} \quad & |B(\ell)| \leq 1 \quad \forall \ell \in \mathcal{L}_{det}. \end{aligned}$$

With error ϵ, $\nu_\epsilon(\mathbf{p}) := \min_{\mathbf{p}' \in \mathcal{NS}: |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon} \nu(\mathbf{p}')$. We call any Bell functional that satisfies the constraint in the above linear program *normalized Bell functional*.

In this definition and in the rest of the chapter, unless otherwise specified (in particular in Lemma 5.2.5), $B(\mathbf{p})$ denotes $\sum_{a,b,x,y} B_{a,b,x,y} p(a,b|x,y)$, where a, b ranges over the non-abort outputs and x, y ranges over the inputs. So even when B and $\mathbf{p}$ have coefficients on the abort events, we do not count them. Table 5.2 summarizes the known upper and lower bounds on ν for various parameters.

Parameter	Upper bound	Ad hoc lower bounds	Best possible lower bound from [BCG$^+$16]
Number of inputs N	$2^c \leq N$ [LS09] [DKLR11, JPPG$^+$10b]	$\frac{\sqrt{N}}{\log(N)}$ [JP11]	$\frac{\sqrt{c}}{q} \leq \log(N)$
Number of outputs K	$O(K)$ [JP11]	$\Omega\left(\frac{K}{(\log(K))^2}\right)$ [BRSdW12]	$\leq \log(K)$
Dimension d	$O(d)$ [JPPG$^+$10b]	$\Omega\left(\frac{d}{(\log(d))^2}\right)$ [BRSdW12]	$\leq \log\log(d)$

Tab. 5.2: Bounds on quantum violations of bipartite normalized Bell inequalities, in terms of the dimension d of the local Hilbert space, the number of settings (or inputs) N and the number of outcomes K (or outputs) per party. In the last column, we compare ad hoc results to the recent constructions of [BCG$^+$16] (Theorem 5.3.1) which gives a lower bound of $\frac{\sqrt{c}}{q}$, where c (resp. q) stands for the classical (resp. quantum) communication complexity of simulating a distribution. We give upper bounds on their construction in terms of the parameters d, N, K.

The (log of the) nuclear norm is a lower bound on classical communication complexity.

Proposition 5.1.2 ([LS09, DKLR11])**.** *For any non-signaling distribution* $\mathbf{p} \in \mathcal{C}$*,* $\mathrm{R}_\epsilon(\mathbf{p})+1 \geq \log(\nu_\epsilon(\mathbf{p}))$*, and for any boolean function* f*,* $\mathrm{R}_\epsilon(f) \geq \log(\nu_\epsilon(\mathbf{p}_f))$*.*

As lower bounds on communication complexity of Boolean functions go, ν is one of the weaker bounds, equivalent to the smooth discrepancy [JK10], and no larger than the approximate nonnegative rank and the smooth rectangle bounds [KMSY14]. More significantly for this work, up to small multiplicative constants, for boolean functions, (the log of) ν is a lower bound on quantum communication, so it is useless to establish gaps between classical and quantum communication complexity. (This limitation, with the upper bound in terms of the number of outputs on normalized Bell violations, is a consequence of Grothendieck's theorem [Gro53].)

The classical and quantum efficiency measures, given here in their dual formulations, are expressed by the following two convex optimization programs. The classical bound is a

generalization to distributions of the partition bound of communication complexity [JK10, LLR12]. This bound is one of the strongest lower bounds known, and can be exponentially larger than ν (an example is the Vector in Subspace problem). It is always as least as large as the relaxed partition bound which is in turn always at least as large as the smooth rectangle bound [JK10, KLL$^+$15]. Its weaker variants have been used to show exponential gaps between classical and quantum communication complexity.

Definition 5.1.3. The *ϵ-error efficiency bound* of a distribution $\mathbf{p}$ is given by

$$\begin{aligned} \mathbf{eff}_\epsilon(\mathbf{p}) := \max_{B,\beta} \quad & \beta \\ \text{subject to} \quad & B(\mathbf{p}') \geq \beta \qquad \forall \mathbf{p}' \text{ s.t. } |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon, \\ & B(\ell) \leq 1 \qquad \forall \ell \in \mathcal{L}_{det}^{\perp}. \end{aligned}$$

We call any Bell functional that satisfies the second constraint in the above linear program *inefficiency-resistant Bell functional*.

In [LLR12], the zero-error efficiency bound was defined in its primal and dual forms as follows

Definition 5.1.4 ([LLR12])**.** The efficiency bound of a distribution $\mathbf{p}$ is given by

$$\begin{aligned} \mathbf{eff}(\mathbf{p}) := \min_{\zeta,\mu_\ell \geq 0} \quad & \frac{1}{\zeta} \\ \text{subject to} \quad & \sum_{\ell \in \mathcal{L}_{det}^{\perp}} \mu_\ell \ell(a,b|x,y) = \zeta p(a,b|x,y) \qquad \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\ & \sum_{\ell \in \mathcal{L}_{det}^{\perp}} \mu_\ell = 1 \\ = \max_{B} \quad & B(\mathbf{p}) \\ \text{subject to} \quad & B(\ell) \leq 1 \quad \forall \ell \in \mathcal{L}_{det}^{\perp} \end{aligned}$$

The ϵ-error efficiency bound was in turn defined as $\min_{\mathbf{p}' | \mathbf{p}' - \mathbf{p}|_1 \leq \epsilon} \mathbf{eff}(\mathbf{p}')$. In the following, we show that this is equivalent to (Definition 5.1.3). In the original definition, the Bell functional could depend on the particular $\mathbf{p}'$. We show that it is always possible to satisfy the constraint with the same Bell functional for all $\mathbf{p}'$ close to $\mathbf{p}$.

In order to prove this, we will need the following notions.

Definition 5.1.5. A *distribution error* $\mathbf{\Delta}$ is a family of additive error terms $\Delta(a,b|x,y) \in [-1,1]$ for all $(a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y}$ such that

$$\sum_{a,b} \Delta(a,b|x,y) = 0 \qquad \forall (x,y) \in \mathcal{X}\times\mathcal{Y}.$$

For any $0 \leq \epsilon \leq 1$, the set Δ_ϵ is the set of distribution errors $\mathbf{\Delta}$ such that

$$\sum_{a,b} |\Delta(a,b|x,y)| \leq \epsilon \qquad \forall (x,y) \in \mathcal{X}\times\mathcal{Y}.$$

This set is a polytope, so it admits a finite set of extremal points. We denote this set by Δ_ϵ^{ext}.

We will use the following properties of Δ_ε.

Fact 5.1.6. For any distribution $\mathbf{p}$, we have

$$\{\mathbf{p}' |\ |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon\} \subseteq \{\mathbf{p} + \mathbf{\Delta} |\ \mathbf{\Delta} \in \Delta_\varepsilon\}.$$

The reason why the set on the right hand side might be larger is that $\mathbf{p} + \mathbf{\Delta}$ might not be a valid distribution. In order to ensure that this is the case, it is sufficient to impose that all obtained purposed probabilities are nonnegative, leading to the following property.

Fact 5.1.7. For any distribution $\mathbf{p}$, we have

$$\{\mathbf{p}' |\ |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon\} = \{\mathbf{p} + \mathbf{\Delta} |\ \mathbf{\Delta} \in \Delta_\varepsilon \ \&\ p(a,b|x,y) + \Delta(a,b|x,y) \geq 0\ \forall a,b,x,y\}.$$

We are now ready to prove the following theorem.

Theorem 5.1.8. *Let* $\mathbf{p}$ *be a distribution,* $\mathbf{eff}_\varepsilon(\mathbf{p})$ *be defined as in Definition 5.1.3 and* $\mathbf{eff}(\mathbf{p})$ *be defined as in Definition 5.1.4. Then, we have*

$$\mathbf{eff}_\varepsilon(\mathbf{p}) = \min_{\mathbf{p}':|\mathbf{p}'-\mathbf{p}|_1\leq\epsilon} \mathbf{eff}(\mathbf{p}').$$

Proof. Let $\overline{\mathbf{eff}}_\varepsilon(\mathbf{p}) = \min_{\mathbf{p}':|\mathbf{p}'-\mathbf{p}|_1\leq\epsilon} \mathbf{eff}(\mathbf{p}')$. We first show that $\mathbf{eff}_\varepsilon(\mathbf{p}) \leq \overline{\mathbf{eff}}_\varepsilon(\mathbf{p})$. Let (B,β) be an optimal feasible point for $\mathbf{eff}_\varepsilon(\mathbf{p})$, so that

$$\begin{aligned} &\mathbf{eff}_\varepsilon(\mathbf{p}) = \beta, \\ &B(\mathbf{p}') \geq \beta && \forall \mathbf{p}' \text{ s.t. } |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon, \\ &B(\ell) \leq 1 && \forall \ell \in \mathcal{L}_{det}^\perp. \end{aligned}$$

Therefore (B,β) is also a feasible point for $\mathbf{eff}(\mathbf{p}')$ for all $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon$, so that $\mathbf{eff}(\mathbf{p}') \geq \beta$ for all such $\mathbf{p}'$, and $\overline{\mathbf{eff}}_\varepsilon(\mathbf{p}) \geq \beta = \mathbf{eff}_\varepsilon(\mathbf{p})$.

It remains to show that $\mathbf{eff}_\varepsilon(\mathbf{p}) \geq \overline{\mathbf{eff}}_\varepsilon(\mathbf{p})$. In order to do so, we first use the primal form of $\mathbf{eff}(\mathbf{p}')$ in Definition 5.1.4 to express $\overline{\mathbf{eff}}_\varepsilon(\mathbf{p})$ as follows:

$$\begin{aligned} \overline{\mathbf{eff}}_\varepsilon(\mathbf{p}) = \min_{\mathbf{p}':|\mathbf{p}'-\mathbf{p}|_1\leq\epsilon} \quad & \mathbf{eff}(\mathbf{p}') \\ = \min_{\zeta,\mu_\ell\geq 0,\mathbf{p}'} \quad & \frac{1}{\zeta} \\ \text{subject to} \quad & \sum_{\ell\in\mathcal{L}_{det}^\perp} \mu_\ell \ell(a,b|x,y) = \zeta p'(a,b|x,y) \quad \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\ & \sum_{\ell\in\mathcal{L}_{det}^\perp} \mu_\ell = 1, \quad |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon \\ = \min_{\zeta,\mu_\ell\geq 0,\mathbf{\Delta}\in\Delta_\varepsilon} \quad & \frac{1}{\zeta} \\ \text{subject to} \quad & \sum_{\ell\in\mathcal{L}_{det}^\perp} \mu_\ell \ell(a,b|x,y) = \\ & \qquad \zeta[p(a,b|x,y) + \Delta(a,b|x,y)] \qquad \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\ & \sum_{\ell\in\mathcal{L}_{det}^\perp} \mu_\ell = 1, \end{aligned}$$

where the last equality follows from Fact 5.1.7 and the fact that the first condition of the program imposes that $p(a,b|x,y)+\Delta(a,b|x,y)$ is nonnegative (since $\sum_\ell \mu_\ell \ell(a,b|x,y)$ is nonnegative). Since Δ_ε is a polytope, $\overline{\mathbf{eff}}_\varepsilon(\mathbf{p})$ can be expressed as the following linear program:

$$\begin{aligned}
\overline{\mathbf{eff}}_\varepsilon(\mathbf{p}) = \min_{\zeta,\mu_\ell\geq 0,\nu_\Delta\geq 0} \quad & \frac{1}{\zeta} \\
\text{subject to} \quad & \sum_{\ell\in\mathcal{L}_{det}^\perp} \mu_\ell \ell(a,b|x,y) = \zeta[p(a,b|x,y)+ \\
& \qquad \sum_{\mathbf{\Delta}\in\Delta_\varepsilon^{ext}} \nu_\Delta \Delta(a,b|x,y)] \qquad\qquad \forall (a,b,x,y)\in\mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\
& \sum_{\ell\in\mathcal{L}_{det}^\perp} \mu_\ell = 1, \quad \sum_{\mathbf{\Delta}\in\Delta_\varepsilon^{ext}} \nu_\Delta = 1.
\end{aligned}$$

Note that this can be written in standard LP form via the change of variables $\mu_\ell = \zeta w_\ell$. By LP duality, we then obtain:

$$\begin{aligned}
\overline{\mathbf{eff}}_\varepsilon(\mathbf{p}) = \max_{B,\beta} \quad & \beta \\
\text{subject to} \quad & B(\mathbf{p}+\mathbf{\Delta}) \geq \beta \qquad \forall \mathbf{\Delta}\in\Delta_\varepsilon, \\
& B(\ell) \leq 1 \qquad\qquad \forall \ell\in\mathcal{L}_{det}^\perp.
\end{aligned}$$

Comparing this to the definition of $\mathbf{eff}_\varepsilon(\mathbf{p})$ (Definition 5.1.3) and together with Fact 5.1.6, we therefore have $\overline{\mathbf{eff}}_\varepsilon(\mathbf{p}) \leq \mathbf{eff}_\varepsilon(\mathbf{p})$.

□

The ϵ-error quantum efficiency bound of a $\mathbf{p}$ is

$$\begin{aligned}
\mathbf{eff}^*_\epsilon(\mathbf{p}) = \max_{B,\beta} \quad & \beta \\
\text{subject to} \quad & B(\mathbf{p}') \geq \beta \qquad \forall \mathbf{p}' \text{ s.t. } |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon, \\
& B(\mathbf{q}) \leq 1 \qquad\qquad \forall \mathbf{q}\in\mathcal{Q}^\perp.
\end{aligned}$$

We denote $\mathbf{eff} = \mathbf{eff}_0$ and $\mathbf{eff}^* = \mathbf{eff}^*_0$ the 0-error bounds.

For any given distribution $\mathbf{p}$, its classical communication complexity is bounded below by the (log of the) efficiency. For randomized communication complexity with error ϵ, the bound is $\log(\mathbf{eff}_\epsilon)$ and for quantum communication complexity, the bound is $\log(\mathbf{eff}^*_\epsilon)$. Note that for any $\mathbf{p}\in\mathcal{Q}$, the quantum communication complexity is 0 and the $\mathbf{eff}^*$ bound is 1. For any function f, the efficiency bound $\mathbf{eff}_\epsilon(\mathbf{p}_f)$ is equivalent to the partition bound [JK10, LLR12].

Proposition 5.1.9 ([LLR12])**.** *For any $\mathbf{p}\in\mathcal{P}$ and any $0\leq\epsilon<1/2$, $\mathrm{R}_\epsilon(\mathbf{p}) \geq \log(\mathbf{eff}_\epsilon(\mathbf{p}))$ and $Q_\epsilon(\mathbf{p}) \geq \frac{1}{2}\log(\mathbf{eff}^*_\epsilon(\mathbf{p}))$. For any $\mathbf{p}\in\mathcal{C}$ and any $0\leq\epsilon\leq 1$, $\nu_\epsilon(\mathbf{p}) \leq 2\mathbf{eff}_\epsilon(\mathbf{p})$.*

Theorem 5.3.2 below involves upper bounds on the quantum efficiency bound. To give an upper bound on the quantum efficiency of a distribution $\mathbf{p}$, it is more convenient to use the primal formulation, and upper bounds can be given by exhibiting a local (or quantum) distribution with abort which satisfies the following two properties: the probability of aborting should be the same on all inputs x, y, and conditioned on not aborting, the outputs of the protocol should reproduce the distribution $\mathbf{p}$. The efficiency is inversely proportional to the probability of not aborting, so the goal is to abort as little as possible.

Proposition 5.1.10 ([LLR12]). *For any distribution* $\mathbf{p}$, $\mathbf{eff}^*(\mathbf{p}) = 1/\eta^*$, *with* η^* *the optimal value of the following optimization problem (non-linear, because* $\mathcal{Q}^\perp$ *is not a polytope).*

$$\begin{aligned} \max_{\zeta, \mathbf{q} \in \mathcal{Q}^\perp} \quad & \zeta \\ \textit{subject to} \quad & q(a,b|x,y) = \zeta p(a,b|x,y) \quad \forall x,y,a,b \in \mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B} \end{aligned}$$

Moreover, for any $0 \leq \epsilon \leq 1$, $\mathbf{eff}^*_\epsilon(\mathbf{p}) = \min_{\mathbf{p}':|\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon} \mathbf{eff}^*(\mathbf{p}')$.

5.2 Properties of Bell inequalities

Syntactically, there are two differences between the normalized Bell functionals (Definition 5.1.1) and the inefficiency-resistant ones (Definition 5.1.3). The first difference is that the normalization constraint is relaxed: for inefficiency-resistant functionals, the lower bound on the Bell value for local distributions is removed. Since this is a maximization problem, this relaxation allows for larger violations.

This difference alone would not lead to a satisfactory measure of nonlocality, since one could obtain unbounded violations by shifting and dilating the Bell functional. The second difference prevents this. The upper bound is required to hold not only for local distributions, but also those that can abort. This is a much stronger condition. Notice that a local distribution can selectively abort on configurations that would otherwise tend to keep the Bell value small, making it harder to satisfy the constraint.

In this section, we show that normalized Bell violations can be modified to be resistant to local distributions that abort, while preserving the violation on any non-signaling distribution, up to a factor of 3. This means that we can add the stronger constraint of resistance to local distributions that abort to Definition 5.1.1, incurring a loss of just a factor of 3, and the only remaining difference between the resulting linear programs is the relaxation of the lower bound (dropping the absolute value) for local distributions that abort.

Theorem 5.2.1. *Let* B *be a normalized Bell functional on* $\mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y}$ *and* $\mathbf{p} \in \mathcal{C}$ *a non-signaling distribution such that* $B(\mathbf{p}) \geq 1$. *Then there exists a normalized Bell functional* B^* *on* $(\mathcal{A}\cup\{\perp\})\times(\mathcal{B}\cup\{\perp\})\times\mathcal{X}\times\mathcal{Y}$ *with* 0 *coefficients on the* $\perp$ *outputs such that :*

$$\begin{aligned} B^*(\mathbf{p}) &\geq \frac{1}{3}B(\mathbf{p}) - \frac{2}{3}, \qquad \forall \mathbf{p} \in \mathcal{NS} \\ |B^*(\ell)| &\leq 1, \qquad\qquad\quad \forall \ell \in \mathcal{L}^\perp_{\text{det}} \end{aligned}$$

The rest of this section is devoted to the proof of Theorem 5.2.1. First, we show (see Observation 5.2.2) how to rescale a normalized Bell functional so that it saturates its normalization constraint. Then, Definition 5.2.3 adds weights to abort events to make the Bell functional resistant to inefficiency. Finally, Lemma 5.2.5 removes the weights on the abort events of a Bell functional while keeping it bounded on the local set with abort, without dramatically changing the values it takes on the non-signaling set. Our techniques are similar to the ones used in [MPRG02].

Observation 5.2.2. *Let* B *be a non-constant normalized Bell functional and* $\mathbf{p} \in \mathcal{C}$ *such that* $B(\mathbf{p}) \geq 1$. *Consider* $\ell^- \in \mathcal{L}^\perp_{\text{det}}$ *such that* $B(\ell^-) = m = \min\{B(\ell)|\ell \in \mathcal{L}^\perp_{\text{det}}\}$ *and* $\ell^+ \in \mathcal{L}^\perp_{\text{det}}$ *such*

that $B(\ell^+) = M = \max\{B(\ell)|\ell \in \mathcal{L}^\perp_{\text{det}}\}$. *We have* $m < M$ *because* B *is non-constant. The Bell functional* $\tilde{B}$ *defined by* $\tilde{B} := \frac{1}{M-m}(2B - M - m)$, *is such that* $\tilde{B}(\ell^+) = 1$, $\tilde{B}(\ell^-) = -1$, $|\tilde{B}(\ell)| \leq 1$ *for all* $\ell \in \mathcal{L}^\perp_{\text{det}}$, *and* $\tilde{B}(\mathbf{p}) \geq B(\mathbf{p})$ *since* B *is normalized.*

Definition 5.2.3 below is the first step of the construction. It takes two marginal distributions $\mathbf{m}_A$ and $\mathbf{m}_B$, and a normalized Bell functional B, and constructs a Bell functional $B^\perp_{\mathbf{m}_A,\mathbf{m}_B}$ whose value over every distribution $\mathbf{p} \in \mathcal{NS}^\perp$ coincides with the value of B over the distribution $\mathbf{p}' \in \mathcal{NS}$ obtained from $\mathbf{p}$ by replacing the abort events with samples from $\mathbf{m}_A$ and $\mathbf{m}_B$.

Definition 5.2.3. For all two families of distributions, $\mathbf{m}_A = (m_A(\cdot|x))_{x\in\mathcal{X}}$ over outcomes in $\mathcal{A}$ for Alice and $\mathbf{m}_B = (m_B(\cdot|y))_{y\in\mathcal{Y}}$ over outcomes in $\mathcal{B}$ for Bob, and any normalized Bell functional B with coefficients only on non-abort events, we define the Bell functional $B^\perp_{\mathbf{m}_A,\mathbf{m}_B}$ on $(\mathcal{A}\cup\{\perp\}) \times (\mathcal{B}\cup\{\perp\}) \times \mathcal{X} \times \mathcal{Y}$ by

$$(B^\perp_{\mathbf{m}_A,\mathbf{m}_B})_{a,b,x,y} := B_{a,b,x,y} + \delta_{a=\perp} \sum_{a'\neq\perp} m_A(a'|x) B_{a',b,x,y} + \delta_{b=\perp} \sum_{b'\neq\perp} m_B(b'|y) B_{a,b',x,y} + \delta_{a=\perp}\delta_{b=\perp} \sum_{a',b'\neq\perp} m_A(a'|x) m_B(b'|y) B_{a',b',x,y}$$

Observation 5.2.4. *Let* $f_{\mathbf{m}_A,\mathbf{m}_B} : \mathcal{NS}^\perp \to \mathcal{NS}$ *be the function that replaces abort events on Alice's (resp. Bob's) side by a sample from* $\mathbf{m}_A$ *(resp.* $\mathbf{m}_B$*) (note that* $f_{\mathbf{m}_A,\mathbf{m}_B}$ *preserves locality). Then, for every* $\mathbf{m}_A$, $\mathbf{m}_B$ *and* B *as in Definition 5.2.3, the Bell functional* $B^\perp_{\mathbf{m}_A,\mathbf{m}_B}$ *satisfies that* $B^\perp_{\mathbf{m}_A,\mathbf{m}_B}(\mathbf{p}) = B(f_{\mathbf{m}_A,\mathbf{m}_B}(\mathbf{p}))$, $\forall \mathbf{p} \in \mathcal{NS}^\perp$, *so* $B^\perp_{\mathbf{m}_A,\mathbf{m}_B}(\mathbf{p}) = B(\mathbf{p})$, *for all* $\mathbf{p} \in \mathcal{NS}$, *and* $|B^\perp_{\mathbf{m}_A,\mathbf{m}_B}(\ell)| \leq 1$, *for all* $\ell \in \mathcal{L}^\perp$.

Next, in Lemma 5.2.5 below, we do without the abort coefficients in the Bell functionals $B^\perp_{\mathbf{m}_A,\mathbf{m}_B}$.

Lemma 5.2.5. *Let* B' *be a normalized Bell functional on* $\mathcal{A}^\perp \times \mathcal{B}^\perp \times \mathcal{X} \times \mathcal{Y}$ *(possibly with non-zero weights on* $\perp$*). Then the Bell functional* B'' *on the same set defined by*

$$B''_{a,b,x,y} = B'_{a,b,x,y} - B'_{a,\perp,x,y} - B'_{\perp,b,x,y} + B'_{\perp,\perp,x,y}, \tag{5.1}$$

for all $(a,b,x,y) \in (\mathcal{A}\cup\{\perp\}) \times (\mathcal{B}\cup\{\perp\}) \times \mathcal{X} \times \mathcal{Y}$ *satisfies :*

1. *If* $a = \perp$ *or* $b = \perp$ *then* $B''_{a,b,x,y} = 0$
2. *for all* $\mathbf{p} \in \mathcal{NS}^\perp$,

$$B''(\mathbf{p}) = B'(\mathbf{p}) - B'(\mathbf{p}_{A,\perp}) - B'(\mathbf{p}_{\perp,B}) + B'(\mathbf{p}_{\perp,\perp}), \tag{5.2}$$

where $\mathbf{p}_{A,\perp} \in \mathcal{L}^\perp$ *(resp.* $\mathbf{p}_{\perp,B} \in \mathcal{L}^\perp$*) is the local distribution obtained from* $\mathbf{p}$ *if Bob (resp. Alice) replaces all of his (resp. her) outputs by* $\perp$, *and* $\mathbf{p}_{\perp,\perp} \in \mathcal{L}^\perp$ *is the local distribution where both Alice and Bob always output* $\perp$. *In Item 2 above, for all* $\mathbf{p}'$,

$$B'(\mathbf{p}') = \sum_{(a,b)\in\mathcal{A}^\perp\times\mathcal{B}^\perp} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} B'_{a,b,x,y} \mathbf{p}'(a,b|x,y)$$

where the first sum is also over the abort events.

Proof. Item 1 follows from (5.1). We prove Item 2. For $\mathbf{p} \in \mathcal{NS}^{\perp}$ with marginals $\mathbf{p}_A$ and $\mathbf{p}_B$, we have: for all $y \in Y$, $p_A(a|x) = \sum_{b \in \mathcal{B}^{\perp}} p(a,b|x,y)$, and for all $x \in X$, $p_B(b|y) = \sum_{a \in \mathcal{A}^{\perp}} p(a,b|x,y)$. For the remainder of this proof, summations involving a (resp. b) are over $a \in \mathcal{A}^{\perp}$ (resp. $b \in \mathcal{B}^{\perp}$).

By definition, $p_{A,\perp}(a,b|x,y) = p_A(a|x)\delta_{b=\perp}$, $p_{\perp,B}(a,b|x,y) = \delta_{a=\perp}p_B(b|y)$, and $p_{\perp,\perp}(a,b|x,y) = \delta_{a=\perp}\delta_{b=\perp}$. We have:

$$\begin{aligned}
B''(\mathbf{p}) &= \sum_{a,b,x,y} \left[B'_{a,b,x,y} - B'_{a,\perp,x,y} - B'_{\perp,b,x,y} + B'_{\perp,\perp,x,y}\right] p(a,b|x,y) \\
&= \sum_{a,b,x,y} B'_{a,b,x,y} p(a,b|x,y) - \sum_{a,x,y} B'_{a,\perp,x,y} \sum_{b} p(a,b|x,y) \\
&\qquad - \sum_{b,x,y} B'_{\perp,b,x,y} \sum_{a} p(a,b|x,y) + \sum_{x,y} B'_{\perp,\perp,x,y} \sum_{a,b} p(a,b|x,y) \\
&= B'(\mathbf{p}) - \sum_{a,x,y} B'_{a,\perp,x,y} p_A(a|x) - \sum_{b,x,y} B'_{\perp,b,x,y} p_B(b|y) + \sum_{x,y} B'_{\perp,\perp,x,y} \\
&= B'(\mathbf{p}) - B'(\mathbf{p}_{A,\perp}) - B'(\mathbf{p}_{\perp,B}) + B'(\mathbf{p}_{\perp,\perp}).
\end{aligned}$$

□

We are now ready to prove Theorem 5.2.1.

Proof of Theorem 5.2.1. If B is constant, since it is normalized by assumption, we have $B \equiv 1$. Thus, we can simply take B^* defined by: for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, $B^*_{a,b,x,y} = B_{a,b,x,y}$ if $(a,b) \in \mathcal{A} \times \mathcal{B}$, and $B^*_{a,b,x,y} = 0$ otherwise.

Now, let us assume that B is not constant. From Observation 5.2.2, we can assume that there exists $\ell^-, \ell^+ \in \mathcal{L}_{\text{det}}$ such that $B(\ell^-) = -1$ and $B(\ell^+) = 1$ (otherwise, we replace B by its saturated version $\tilde{B}$). Since ℓ^- and ℓ^+ are deterministic distributions, we have: $\ell^- = \ell^-_A \otimes \ell^-_B$ and $\ell^+ = \ell^+_A \otimes \ell^+_B$, for some marginals $\ell^-_A, \ell^-_B, \ell^+_A$, and ℓ^+_B. We consider the replacing Bell functional $B^{\perp}_{\ell^-_A,\ell^-_B}$ (resp. $B^{\perp}_{\ell^+_A,\ell^+_B}$) from Definition 5.2.3 constructed from (B, ℓ^-_A, ℓ^-_B) (resp. from (B, ℓ^+_A, ℓ^+_B)). Taking $B' = \frac{1}{2}(B^{\perp}_{\ell^-_A,\ell^-_B} + B^{\perp}_{\ell^+_A,\ell^+_B})$, we have $|B'(\ell)| \leq 1$, for all $\ell \in \mathcal{L}^{\perp}$, and therefore we can apply Lemma 5.2.5 to get B'' from B'. Since $B'(\mathbf{p}_{\perp,\perp}) = \frac{1}{2}(B^{\perp}_{\ell^-_A,\ell^-_B}(\mathbf{p}_{\perp,\perp}) + B^{\perp}_{\ell^+_A,\ell^+_B}(\mathbf{p}_{\perp,\perp})) = \frac{1}{2}(B(\ell^-) + B(\ell^+)) = 0$, by (5.2) we have for all $\mathbf{p} \in \mathcal{NS}^{\perp}$, $B''(\mathbf{p}) = B'(\mathbf{p}) - B'(\mathbf{p}_{A,\perp}) - B'(\mathbf{p}_{\perp,B})$. Hence, denoting $B^* = \frac{1}{3}B''$, B^* satisfies all the required properties since $|B'(\ell)| \leq 1$ for all $\ell \in \mathcal{L}^{\perp}$ and therefore we have for all $\mathbf{p} \in \mathcal{NS}$, $B^*(\mathbf{p}) \geq \frac{1}{3}B'(\mathbf{p}) - \frac{1}{3}|B'(\mathbf{p}_{A,\perp})| - \frac{1}{3}|B'(\mathbf{p}_{\perp,B})| \geq \frac{1}{3}B'(\mathbf{p}) - \frac{2}{3}$, and for all $\ell \in \mathcal{L}^{\perp}$, $|B^*(\ell)| \leq \frac{1}{3}|B'(\ell)| + \frac{1}{3}|B'(\ell_{A,\perp})| + \frac{1}{3}|B'(\ell_{\perp,B})| \leq 1$. □

5.3 Exponential violations from communication bounds

In a recent paper, Buhrman *et al.* gave a general construction to derive normalized Bell inequalities from any sufficiently large gap between classical and quantum communication complexity.

Theorem 5.3.1 ([BCG+16])**.** *For any function f for which there is a quantum protocol using q qubits of communication but no prior shared entanglement, there exists a quantum*

distribution $\mathbf{q} \in \mathcal{Q}$ *and a normalized Bell functional* B *such that*

$$B(\mathbf{q}) \geq \frac{\sqrt{R_{1/3}(f)}}{6\sqrt{3}q}(1-2^{-q})^{2q}.$$

Their construction is quite involved, requiring protocols to be memoryless, which they show how to achieve in general, and uses port-based teleportation [IH08, IH09] to construct a quantum distribution. The Bell inequality they construct expresses a correctness constraint.

In this section, we show how to obtain large inefficiency-resistant Bell violations for quantum distributions from gaps between quantum communication and classical communication lower bounds. We first prove the stronger of two statements, which gives violations of $\frac{\mathbf{eff}_\epsilon(\mathbf{p})}{\mathbf{eff}^\star_{\epsilon'}(\mathbf{p})}$.For any problem for which a classical lower bound c is given using the efficiency or partition bound or any weaker method (including the rectangle bound and its variants), and any upper bound q on quantum communication complexity, it implies a violation of 2^{c-2q}.

Theorem 5.3.2. *For any distribution* $\mathbf{p}$ *and any* $0 \leq \epsilon' \leq \epsilon \leq 1$*, if* (B, β) *is a feasible solution to the dual of* $\mathbf{eff}_\epsilon(\mathbf{p})$ *and* $(\zeta, \mathbf{q})$ *is a feasible solution to the primal for* $\mathbf{eff}^\star_{\epsilon'}(\mathbf{p})$*, then there is a quantum distribution* $\overline{\mathbf{q}} \in \mathcal{Q}$ *such that*

$$B(\overline{\mathbf{q}}) \geq \zeta\beta \qquad \text{and} \qquad B(\ell) \leq 1, \quad \forall \ell \in \mathcal{L}^\perp_{det}\,,$$

and in particular, if both are optimal solutions, then

$$B(\overline{\mathbf{q}}) \geq \frac{\mathbf{eff}_\epsilon(\mathbf{p})}{\mathbf{eff}^\star_{\epsilon'}(\mathbf{p})}\,.$$

The distribution $\overline{\mathbf{q}}$ *has one additional output per player compared to the distribution* $\mathbf{p}$*.*

Proof. Let (B, β) be a feasible solution to the dual of $\mathbf{eff}_\epsilon(\mathbf{p})$, $\mathbf{p}'$ be such that $\mathbf{eff}^\star_{\epsilon'}(\mathbf{p}) = \mathbf{eff}^\star(\mathbf{p}')$ with $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon'$, and $(\zeta, \mathbf{q})$ be a feasible solution to the primal for $\mathbf{eff}^\star(\mathbf{p}')$. From the constraints, we have

$$\begin{aligned}
&\mathbf{q} \in \mathcal{Q}^\perp, \\
&q(a,b|x,y) = \zeta p'(a,b|x,y) && \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y}, \\
&B(\ell) \leq 1 && \forall \ell \in \mathcal{L}^\perp_{det}, \\
&B(\mathbf{p}'') \geq \beta && \forall \mathbf{p}'' \text{ s.t. } |\mathbf{p}'' - \mathbf{p}|_1 \leq \epsilon.
\end{aligned}$$

Then $B(\mathbf{q}) = \zeta B(\mathbf{p}') \geq \zeta\beta$. However, $\mathbf{q} \in \mathcal{Q}^\perp$ but technically we want a distribution in $\mathcal{Q}$ (not one that aborts). So we add a new (valid) output 'A' to the set of outputs of each player, and they should output 'A' instead of aborting whenever $\mathbf{q}$ aborts. The resulting distribution, say $\overline{\mathbf{q}} \in \mathcal{Q}$ (with additional outcomes 'A' on both sides), is such that $B(\overline{\mathbf{q}}) = B(\mathbf{q})$ (since the Bell functional B does not have any weight on $\perp$ or on 'A'). □

Theorem 5.3.1 [BCG+16] and Theorem 5.3.2 are both general constructions, but there are a few significant differences worth pointing out. Firstly, our Theorem 5.3.2 requires a lower bound on the partition bound in the numerator, whereas Theorem 5.3.1 only requires a lower bound on communication complexity (which could be exponentially larger). Secondly, Theorem 5.3.1 requires a quantum communication protocol in the denominator, whereas our

theorem only requires an upper bound on the quantum efficiency (which could be exponentially smaller). Thirdly, although our bound is exponentially larger than Buhrman *et al.*'s for most problems considered here, and applies to subquadratic gaps, their bounds are of the more restricted class of normalized Bell inequalities.

Theorem 5.3.2 gives an explicit Bell functional B provided an explicit solution to the efficiency (partition) bound is given and the quantum distribution is obtained from a solution to the primal of $\mathbf{eff}^\star$ (Proposition 5.1.10). Recall that giving a solution to the primal of $\mathbf{eff}^\star$ consists in exhibiting a quantum zero-communication protocol that can abort, which conditioned on not aborting, outputs following $\mathbf{p}$.

We can also start from a quantum protocol, as we show below. From the quantum protocol, we derive a quantum distribution using standard techniques.

Corollary 5.3.3. *For any distribution* $\mathbf{p}$ *and any* $0 \leq \epsilon' \leq \epsilon \leq 1$ *such that* $R_\epsilon(\mathbf{p}) \geq \log(\mathbf{eff}_\epsilon(\mathbf{p})) \geq c$ *and* $Q_{\epsilon'}(\mathbf{p}) \leq q$*, there exists an explicit inefficiency-resistant* B *derived from the efficiency lower bound, and an explicit quantum distribution* $\overline{\mathbf{q}} \in \mathcal{Q}$ *derived from the quantum protocol such that* $B(\overline{\mathbf{q}}) \geq 2^{c-2q}$*.*

Proof. Let (B, β) be an optimal solution to $\mathbf{eff}_\epsilon(\mathbf{p})$ and let c be such that $\mathbf{eff}_\epsilon(\mathbf{p}) = \beta \geq 2^c$. By optimality of B, we have $B(\mathbf{p}') \geq 2^c$ for any $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon$. Since $Q_{\epsilon'}(\mathbf{p}) \leq q$, there exists a q-qbit quantum protocol for some distribution $\mathbf{p}'$ with $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon$. Then, we can use teleportation to obtain a $2q$ classical bit, entanglement-assisted protocol for $\mathbf{p}'$. We can simulate it without communication by picking a shared $2q$-bit random string and running the protocol but without sending any messages. If the measurements do not match the string, output a new symbol 'A' (not in the output set of the quantum protocol and different from $\perp$). We obtain a quantum distribution $\overline{\mathbf{q}}$ such that $B(\overline{\mathbf{q}}) = B(\mathbf{p}')/2^{2q} \geq 2^{c-2q}$. □

Most often, communication lower bounds are not given as efficiency or partition bounds, but rather using variants of the corruption bound. We show in Section 5.5.1 how to map a corruption bound to explicit Bell coefficients.

5.4 Noise-resistant violations from communication bounds

Normalized Bell inequalities are naturally resistant to any local noise: if the observed distribution is $\tilde{\mathbf{p}} = (1-\varepsilon)\mathbf{p} + \varepsilon\ell$ for some $\ell \in \mathcal{L}$, then $B(\tilde{\mathbf{p}}) \geq (1-\varepsilon)B(\mathbf{p}) - \varepsilon$ since $|\, B(\ell)\,| \leq 1$. In inefficiency-resistant Bell inequalities, relaxing the absolute value leads to the possibility that $B(\ell)$ has a large negative value for some local ℓ. (Indeed, such large negative values are inherent to large gaps between ν and $\mathbf{eff}$.) If this distribution were to be used as adversarial noise, then the observed distribution, $(1-\varepsilon)\mathbf{p} + \varepsilon\ell$, would have a Bell value that could be much less than 1. This makes inefficiency-resistant Bell inequalities susceptible to adversarial local noise.

Our construction from Theorem 5.3.2 is susceptible to uniform noise since most of the time, the output is 'A'. Uniform noise will disproportionately hit the non-'A' outputs, destroying the structure of the distribution. In Theorem 5.4.1, we show that our construction can be made resistant to uniform noise, by including a (possible) transcript from the protocol in the outputs. (Notice that this leads to a much larger output set.) Since the transcripts in our construction are teleportation measurements, they follow a uniform distribution, making the modified distribution resistant to uniform noise. The tolerance to noise comes from the error parameter in the classical communication lower bound.

Let $N_\varepsilon(\mathbf{p})$ be the ε-noisy neighbourhood of $\mathbf{p}$, defined as

$$N_\varepsilon(\mathbf{p}) = \{(1-\delta)\mathbf{p} + \delta\mathbf{u} \mid \delta \in [0,\varepsilon]\}$$

where $\mathbf{u}$ the uniform noise distribution, that is: $u(a,b|x,y) = \frac{1}{|\mathcal{A}|\cdot|\mathcal{B}|}$ for all $(a,b) \in \mathcal{A}\times\mathcal{B}$.

Theorem 5.4.1. *For any distribution* $\mathbf{p}$ *and any* $0 \leq \epsilon' \leq \epsilon \leq 1$ *such that* $R_\epsilon(\mathbf{p}) \geq \log(\mathbf{eff}_\epsilon(\mathbf{p})) \geq c$ *and* $Q_{\epsilon'}(\mathbf{p}) \leq q$*, there exists an explicit inefficiency-resistant* $\tilde{B}$ *derived from the efficiency lower bound, and an explicit quantum distribution* $\overline{\mathbf{q}} \in \mathcal{Q}$ *derived from the quantum protocol such that* $\tilde{B}(\mathbf{q}') \geq 2^{c-2q}$ *for any* $\mathbf{q}' \in N_{\varepsilon-\varepsilon'}(\overline{\mathbf{q}})$.

Proof. From a quantum communication protocol for $\mathbf{p}'$ with $|\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon'$ using q qubits of communication, we construct an entanglement-assisted protocol using $2q$ bits of communication and teleportation. Let $\mathcal{M}_A$ (resp. $\mathcal{M}_B$) be the set of possible transcripts for Alice (resp. Bob), with $|\mathcal{M}_A| = M_A$ (resp. $|\mathcal{M}_B| = M_B$), and note that $\log M_A + \log M_B = 2q$

We define the quantum distribution $\overline{\mathbf{q}}$ where Alice's possible outputs are $\mathcal{A}\times\mathcal{M}_A$ and Bob's possible outputs are $\mathcal{B}\times\mathcal{M}_B$. Alice proceeds as follows (Bob proceeds similarly):

1. She runs the quantum protocol for $\mathbf{p}'$ as if all bits received from Bob were 0.
2. She outputs (a,μ_A), where μ_A is the transcript of the messages she would have sent to Bob and a is the output she would have produced in the original protocol.

By definition, this distribution is such that, for all a,b,x,y, $\overline{q}(a,0,b,0|x,y) = \frac{1}{2^{2q}}p'(a,b|x,y)$.

Let $\mathbf{eff}_\varepsilon(\mathbf{p}) \geq 2^c$ be achieved by the Bell functional B. By definition, we have

$$\begin{aligned} B(\ell) &\leq 1 \qquad && \forall \ell \in \mathcal{L}_{det}^{\perp} \\ B(\mathbf{p}'') &\geq 2^c \qquad && \forall \mathbf{p}'' \text{ such that } |\mathbf{p}''-\mathbf{p}|_1 \leq \varepsilon, \end{aligned}$$

In particular for any $\mathbf{p}'' \in N_{\varepsilon-\varepsilon'}(\mathbf{p})$, that is, $\mathbf{p}'' = (1-\delta)\mathbf{p} + \delta\mathbf{u}$ for some $\delta \in [0,\varepsilon-\varepsilon']$, we have $|\mathbf{p}''-\mathbf{p}|_1 \leq \varepsilon$ and therefore

$$B(\mathbf{p}'') = (1-\delta)B(\mathbf{p}') + \delta B(\mathbf{u}) \geq 2^c,$$

where $B(\mathbf{u}) = \frac{1}{|\mathcal{A}\times\mathcal{B}|}\sum_{a,b,x,y} B_{a,b,x,y}$.

Let the Bell functional $\tilde{B}$ for distributions over $(\mathcal{A}\times\mathcal{M}_A)\times(\mathcal{B}\times\mathcal{M}_B)$ be defined as follows

$$\tilde{B}_{(a,\mu_A),(b,\mu_B),x,y} = \begin{cases} B_{a,b,x,y} & \text{if } \mu_A = \mu_B = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Let $\tilde{\mathcal{L}}_{det}^{\perp}$ be the local set for distributions over $(\mathcal{A}\times\mathcal{M}_A)\times(\mathcal{B}\times\mathcal{M}_B)$. The new Bell functional satisfies $\tilde{B}(\ell) \leq 1$ for all $\ell \in \tilde{\mathcal{L}}_{det}^{\perp}$ (by assimilating any event with $\mu_A \neq 0$ or $\mu_B \neq 0$ to an abort event), as well as $\tilde{B}(\overline{\mathbf{q}}) = \frac{1}{2^{2q}}B(\mathbf{p}')$. Therefore, $\forall\delta \in [0,\varepsilon-\varepsilon']$, we also have

$$\begin{aligned} (1-\delta)\tilde{B}(\overline{\mathbf{q}}) + \delta\tilde{B}(\mathbf{u}) &= (1-\delta)\frac{1}{2^{2q}}B(\mathbf{p}') + \delta\frac{1}{|\mathcal{A}\times\mathcal{B}|M_AM_B}\sum_{a,\mu_A,b,\mu_B,x,y}\tilde{B}_{(a,\mu_A),(b,\mu_B),x,y} \\ &= \frac{1}{2^{2q}}\left[(1-\delta)B(\mathbf{p}') + \delta\frac{1}{|\mathcal{A}\times\mathcal{B}|}\sum_{a,b,x,y}B_{a,b,x,y}\right] \\ &= \frac{1}{2^{2q}}\big[(1-\delta)B(\mathbf{p}') + \delta B(\mathbf{u})\big] \geq \frac{2^c}{2^{2q}}. \end{aligned}$$

Therefore, $\forall\mathbf{q}' \in N_{\varepsilon-\varepsilon'}(\overline{\mathbf{q}})$, $\tilde{B}(\mathbf{q}') \geq 2^{c-2q}$, as claimed. □

5.5 Explicit constructions

5.5.1 From corruption bound to Bell inequality violation

The corruption bound, introduced by Yao in [Yao83], is a very useful lower bound technique. It has been used for instance in [Raz92] to get a tight $\Omega(n)$ lower bound on the randomized communication complexity of Disjointness (whereas the approximate rank, for example, can only show a lower bound of $\Theta(\sqrt{n})$). We now explain how to construct an explicit Bell inequality violation from the corruption bound.

Definition 5.5.1. A *rectangle* R of the input space $\mathcal{X} \times \mathcal{Y}$ is a subset of that space of the form $R_A \times R_B$ where $R_A \subseteq \mathcal{X}$ and $R_B \subseteq \mathcal{Y}$.

Theorem 5.5.2 (Corruption bound [Yao83, BFS86, KN97])**.** *Let f be a (possibly partial) Boolean function on $\mathcal{X} \times \mathcal{Y}$. Given $\gamma, \delta \in (0,1)$, suppose that there is a distribution μ on $\mathcal{X} \times \mathcal{Y}$ such that for every rectangle $R \subseteq \mathcal{X} \times \mathcal{Y}$*

$$\mu(R \cap f^{-1}(1)) > \gamma \mu(R \cap f^{-1}(0)) - \delta.$$

Then, for every $\epsilon \in (0,1)$, $2^{R_\epsilon(f)} \geq \frac{1}{\delta}\left(\mu(f^{-1}(0)) - \frac{\epsilon}{\gamma}\right)$.

See, e.g., Lemma 3.5 in [BPSW06] for a rigorous treatment. For several problems, such a μ is already known. In Theorem 5.5.3 below we show how to construct a Bell inequality violation from this type of bound.

Theorem 5.5.3. *Let f be a (possibly partial) Boolean function on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$. Fix $z \in \{0,1\}$. Let μ be an input distribution, and $(U_i)_{i \in I}$ (resp. $(V_j)_{j \in J}$) be a family of pairwise nonoverlapping subsets of $f^{-1}(\bar{z})$ (resp. of $f^{-1}(z)$). Assume that there exists $g : \mathbb{N} \to (0, +\infty)$ and real constants $(u_i)_{i \in I}, (v_i)_{i \in I}$ such that, for any rectangle $R \subseteq \mathcal{X} \times \mathcal{Y}$*

$$\sum_{i \in I} u_i \mu(R \cap U_i) \geq \sum_{j \in J} v_j \mu(R \cap V_j) - g(n). \tag{5.3}$$

Then, the Bell functional B given by the following coefficients: for all $(a,b,x,y) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \times \mathcal{Y}$,

$$B_{a,b,x,y} = \begin{cases} 1/2(-u_i g(n)^{-1} \mu(x,y)) & \text{if } (x,y) \in U_i \text{ and } a \oplus b = z, \\ 1/2(v_j g(n)^{-1} \mu(x,y)) & \text{if } (x,y) \in V_j \text{ and } a \oplus b = z, \\ 0 & \text{otherwise.} \end{cases}$$

satisfies

$$B(\ell) \leq 1, \quad \forall \ell \in \mathcal{L}_{det}^{\perp}, \tag{5.4}$$

$$B(\mathbf{p}_f) = \frac{1}{2 \cdot g(n)} \sum_j v_j \mu(V_j) \tag{5.5}$$

and for any $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}_f|_1 \leq \epsilon$:

$$B(\mathbf{p}') \geq \frac{1}{2 \cdot g(n)} \left[\sum_j v_j \mu(V_j) - \epsilon \left(\sum_j |v_j| \mu(V_j) + \sum_i |u_i| \mu(U_i) \right) \right]. \tag{5.6}$$

Proof. Let us first set $B_{z,x,y} = B_{a,b,x,y}$ for all $a \oplus b = z$. Let $\ell \in \mathcal{L}_{det}^{\perp}$. Then, we have: $B(\ell) = \sum_{(x,y)\in R} B_{z,x,y} + \sum_{(x,y)\in S} B_{z,x,y}$, where R and S are the two rectangles where ℓ outputs z. Let us take a rectangle R. Then

$$\sum_{(x,y)\in R} B_{z,x,y} = \frac{1}{2\cdot g(n)} \left(\sum_j v_j \mu(V_j \cap R) - \sum_i u_i \mu(U_i \cap R) \right) \leq 1/2$$

with the inequality following from (5.3). This proves (5.4). Let us now analyse$B(\mathbf{p}_f)$. By linearity of B and the definition of its coefficients, we have:

$$\begin{aligned}
B(\mathbf{p}_f) &= \sum_{a,b,x,y} B_{a,b,x,y} \mathbf{p}_f(a,b|x,y) \\
&= \frac{1}{2} \sum_{(x,y)\in f^{-1}(z),a,b} B_{a,b,x,y} \chi_{\{z\}}(a\oplus b) + \frac{1}{2} \sum_{(x,y)\in f^{-1}(\bar{z}),a,b} B_{a,b,x,y} \chi_{\{\bar{z}\}}(a\oplus b) \\
&= \frac{1}{2} \sum_j \sum_{(x,y)\in V_j} v_j g(n)^{-1} \mu(x,y) \\
&= \frac{1}{2\cdot g(n)} \sum_j v_j \mu(V_j)
\end{aligned}$$

(for the third equality we used the fact that $B_{a,b,x,y} = 0$ when $a \oplus b = \bar{z}$). This proves (5.5). Moreover, for any family of additive error terms $\Delta(a,b|x,y) \in [-1,1]$ such that

$$\sum_{a,b} |\Delta(a,b|x,y)| \leq \epsilon \qquad\qquad \forall x,y \in \mathcal{X}\times\mathcal{Y},$$

denoted collectively as $\mathbf{\Delta}$, we have

$$\begin{aligned}
|B(\mathbf{\Delta})| &= \left| \sum_{a,b,x,y} B_{a,b,x,y} \Delta(a,b|x,y) \right| \\
&= \frac{1}{2\cdot g(n)} \left| \sum_{a,b\,:\,a\oplus b = z} \left[\sum_i \sum_{(x,y)\in U_i} (-u_i) \mu(x,y) \Delta(a,b|x,y) \right.\right. \\
&\qquad \left.\left. + \sum_j \sum_{(x,y)\in V_j} v_j \mu(x,y) \Delta(a,b|x,y) \right] \right| \\
&\leq \frac{1}{2\cdot g(n)} \left[\sum_i \sum_{(x,y)\in U_i} |u_i| \mu(x,y) \left(\sum_{a,b} |\Delta(a,b|x,y)| \right) \right. \\
&\qquad \left. + \sum_j \sum_{(x,y)\in V_j} |v_j| \mu(x,y) \left(\sum_{a,b} |\Delta(a,b|x,y)| \right) \right] \\
&\leq \frac{\epsilon}{2\cdot g(n)} \left[\sum_i |u_i| \mu(U_i) + \sum_j |v_j| \mu(V_j) \right].
\end{aligned}$$

From this calculation and (5.5), we obtain, for $\mathbf{p}' = \mathbf{p}_f + \mathbf{\Delta}$:

$$B(\mathbf{p}') = B(\mathbf{p}_f) + B(\mathbf{\Delta}) \geq \frac{1}{2 \cdot g(n)} \left[\sum_j v_j \mu(V_j) - \epsilon \left(\sum_j |v_j| \mu(V_j) + \sum_i |u_i| \mu(U_i) \right) \right],$$

which proves (5.6).

□

For many other problems in the literature, such as Vector in Subspace and Tribes, stronger variants of the corruption bound are needed to obtain good lower bounds. These stronger variants have been shown to be no stronger than the partition bound (more specifically, the relaxed partition bound) [KLL+15]. The generalization in Theorem 5.5.3 of the hypothesis of Theorem 5.5.2, which the reader might have notice, allow us to construct explicit Bell functionals also for these problems.

5.5.2 Some specific examples

Using Corollary 5.3.3 and the construction to go from a corruption bound (or its variants) to a Bell inequality (Theorem 5.5.3), we give explicit Bell inequalities and violations for several problems studied in the literature. Since our techniques also apply to small gaps, we include problems for which the gap between classical and quantum communication complexity is polynomial.

Vector in Subspace

In the Vector in Subspace Problem $\mathrm{VSP}_{0,n}$, Alice is given an $n/2$ dimensional subspace of an n dimensional space over $\mathbb{R}$, and Bob is given a vector. This is a partial function, and the promise is that either Bob's vector lies in the subspace, in which case the function evaluates to 1, or it lies in the orthogonal subspace, in which case the function evaluates to 0. Note that the input set of $\mathrm{VSP}_{0,n}$ is continuous, but it can be discretized by rounding, which leads to the problem $\widetilde{\mathrm{VSP}}_{\theta,n}$ (see [KR11] for details). Klartag and Regev [KR11] show that the VSP can be solved with an $O(\log n)$ quantum protocol, but the randomized communication complexity of this problem is $\Omega(n^{1/3})$. As shown in [KLL+15], this is also a lower bound on the relaxed partition bound. Hence Corollary 5.3.3 yields the following.

Proposition 5.5.4. *There exists a Bell inequality B and a quantum distribution $\overline{\mathbf{q}}_{VSP} \in \mathcal{Q}$ such that $B\left(\overline{\mathbf{q}}_{VSP}\right) \in 2^{\Omega(n^{1/3}) - O(\log n)}$ and for all $\ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$.*

Note that the result of [KR11] (Lemma 4.3) is not of the form needed to apply Theorem 5.5.3. It is yet possible to obtain an explicit Bell functional following the proof of Lemma 5.1 in [KLL+15].

Disjointness

In the Disjointness problem, the players receive two sets and have to determine whether they are disjoint or not. More formally, for every n, if we denote with $\mathcal{P}([n])$ the power set of $\{1, \ldots, n\}$, the DISJ_n predicate is defined over $\mathcal{X} = \mathcal{Y} = \mathcal{P}([n])$ by

$$\mathrm{DISJ}_n(x, y) = 1 \text{ iff } x \text{ and } y \text{ are disjoint}$$

It is also convenient to see this predicate as defined over length n inputs, where for $x, y \in \{0,1\}^n$,

$$\mathrm{DISJ}_n(x,y) = 1 \text{ iff } |\{i \,:\, x_i = 1 = y_i\}| = 0$$

In [Raz92], Razborov proved the following.

Lemma 5.5.5 ([Raz92])**.** *There exist two distributions μ_0 and μ_1 with $\mathrm{supp}(\mu_0) \subseteq \mathrm{DISJ}_n^{-1}(1)$ and $\mathrm{supp}(\mu_1) \subseteq \mathrm{DISJ}_n^{-1}(0)$, such that: for any rectangle R in the input space,*

$$\mu_1(R) \geq \Omega(\mu_0(R)) - 2^{\Omega(n)}.$$

Following his proof, one can check that we actually have:

$$\mu_1(R) \geq \frac{1}{45}\mu_0(R) - 2^{-\epsilon n + \log_2(2/9)}.$$

So, letting $\mu := (\mu_0 + \mu_1)/2$,

$$\mu(R \cap f^{-1}(0)) \geq \frac{1}{45}\mu(R \cap f^{-1}(1)) - 2^{-\epsilon n + \log_2(4/9)}. \tag{5.7}$$

Remark 5.5.6. Actually, $\mathrm{supp}(\mu_1) = A_1 := \{(x,y) \,:\, |x| = |y| = m, |x \cap y| = 1\} \subseteq \mathrm{DISJ}_n^{-1}(0)$.

Note that by this construction, $\mu(f^{-1}(0)) = \mu(f^{-1}(1)) = 1/2$. Combining (5.7) with Theorem 5.5.3 (with $g(n) = 2^{-\epsilon n + \log_2(4/9)}$), we obtain:

Corollary 5.5.7. *There exists a Bell inequality B satisfying: $\forall \ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$,*

$$B(\mathbf{p}_{\mathrm{DISJ}_n}) = \frac{1}{90} 2^{\epsilon n - \log_2(4/9)},$$

and for any distribution $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}_{\mathrm{DISJ}_n}|_1 \leq \varepsilon$,

$$B(\mathbf{p}') \geq 2^{\epsilon n - \log_2(4/9)} \frac{1 - 46\epsilon}{90}.$$

More precisely, Theorem 5.5.3 gives an explicit construction of such a Bell inequality: we can define B as:

$$B_{a,b,x,y} = \begin{cases} -2^{\epsilon n - \log_2(4/9)}\mu(x,y) & \text{if } \mathrm{DISJ}_n(x,y) = 0 \text{ and } a \oplus b = 1 \\ \frac{1}{45} 2^{\epsilon n - \log_2(4/9)}\mu(x,y) & \text{if } \mathrm{DISJ}_n(x,y) = 1 \text{ and } a \oplus b = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Combining Corollary 5.3.3 together with the fact there is a quantum protocol for DISJ_n using $O(\sqrt{n})$ communication [AA05] we obtain, through Corollary 5.3.3, the following:

Proposition 5.5.8. *There is a quantum distribution $\overline{\mathbf{q}}_{\mathrm{DISJ}} \in \mathcal{Q}$ and an explicit Bell inequality B satisfying: $B(\overline{\mathbf{q}}_{\mathrm{DISJ}}) = 2^{\Omega(n) - O(\sqrt{n})}$, and for all $\ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$.*

Tribes.

Let $n = (2r+1)^2$ with $r \geq 2$ and let $\text{TRIBES}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be defined as:

$$\text{TRIBES}_n(x,y) := \bigwedge_{i=1}^{\sqrt{n}} \left(\bigvee_{j=1}^{\sqrt{n}} (x_{(i-1)\sqrt{n}+j} \text{ and } y_{(i-1)\sqrt{n}+j}) \right).$$

In [HJ13][Sec. 3] the following is proven:

Lemma 5.5.9. *There exists a probability distribution μ on $\{0,1\}^n \times \{0,1\}^n$ for which there exist numbers $\alpha, \lambda, \gamma, \delta > 0$ such that for sufficiently large n and for any rectangle R in the input space:*

$$\gamma\mu(U_1 \cap R) \geq \alpha\mu(V_1 \cap R) - \lambda\mu(V_2 \cap R) - 2^{-\delta n/2+1}$$

where $U_1 = \text{TRIBES}_n^{-1}(0)$, $\{V_1, V_2\}$ forms a partition of $\text{TRIBES}_n^{-1}(1)$ and $\mu(U_1) = 1 - 7\beta^2/16$, $\mu(V_1) = 6\beta^2/16$, $\mu(V_2) = \beta^2/16$ with $\beta = \frac{r+2}{r+1}$.

In [HJ13], the coefficients are $\alpha = 0.99, \lambda = \frac{16}{3(0.99)^2}$ and $\gamma = \frac{16}{(0.99)^2}$ (the authors say these values have not been optimized).

Combining this result with our Theorem 5.5.3 (taking $z = 1, i = 1, j = 2$, U_1, V_1, V_2 as in Lemma 5.5.9, $u_1 = \gamma, v_1 = \alpha, v_2 = -\lambda$, and $g(n) = 2^{-\delta n/2+1}$), we obtain:

Corollary 5.5.10. *There exists a Bell inequality satisfying: $\forall \ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$,*

$$B(\mathbf{p}_{\text{TRIBES}_n}) = 2^{\delta n/2-1}\frac{\beta^2}{16}(6\alpha - \lambda),$$

and for any distribution $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}_{\text{TRIBES}_n}|_1 \leq \varepsilon$,

$$B(\mathbf{p}') \geq 2^{\delta n/2-1}\left[\frac{\beta^2}{16}(6\alpha - \lambda) - \epsilon(\gamma(1 - 7\beta^2/16) + \lambda\beta^2/16 + \alpha 6\beta^2/16)\right].$$

More precisely, Theorem 5.5.3 provides a Bell inequality B yielding this bound, defined as:

$$B_{a,b,x,y} = \begin{cases} -\gamma 2^{\delta n/2-1}\mu(x,y) & \text{if } (x,y) \in U_1 \text{ and } a \oplus b = 1 \\ \alpha 2^{\delta n/2-1}\mu(x,y) & \text{if } (x,y) \in V_1 \text{ and } a \oplus b = 1 \\ -\lambda 2^{\delta n/2-1}\mu(x,y) & \text{if } (x,y) \in V_2 \text{ and } a \oplus b = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Combining Corollary 5.5.10 together with the fact there is a quantum protocol for TRIBES_n using $O(\sqrt{n}(\log n)^2)$ communication [BCW98] we obtain, through Corollary 5.3.3, the following:

Proposition 5.5.11. *There is a quantum distribution $\overline{\mathbf{q}}_{\text{TRIBES}} \in \mathcal{Q}$ and an explicit Bell inequality B satisfying: $B(\overline{\mathbf{q}}_{\text{TRIBES}}) = 2^{\Omega(n)-O(\sqrt{n}(\log n)^2)}$, and for all $\ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$.*

Gap Orthogonality.

The Gap Orthogonality (ORT) problem was introduced by Sherstov as an intermediate step to prove a lower bound for the Gap Hamming Distance (GHD) problem [She12]. We derive an explicit Bell inequality for ORT from Sherstov's lower bound of $\Omega(n)$, shown in [KLL$^+$15] to be a relaxed partition bound. (Applying Corollary 5.3.3 also gives a (non-explicit) violation for GHD.) The quantum upper bound is $O(\sqrt{n}\log n)$ by the general result of [BCW98]. In the ORT problem, the players receive vectors and need to tell whether they are nearly orthogonal or far from orthogonal. More formally, we consider the input space $\{-1,+1\}^n$ (to stick to the usual notations for this problem), and we denote $\langle\cdot,\cdot\rangle$ the scalar product on $\{-1,+1\}^n$. Let $\mathrm{ORT}_n : \{-1,+1\}^n \times \{-1,+1\}^n \to \{-1,+1\}$ be the partial function defined as in [She12] by:

$$\mathrm{ORT}_n(x,y) = \begin{cases} -1 & \text{if } |\langle x,y\rangle| \leq \sqrt{n} \\ +1 & \text{if } |\langle x,y\rangle| \geq 2\sqrt{n}. \end{cases}$$

Let f_n be the partial functions over $\{-1,+1\}^n \times \{-1,+1\}^n$ by $f_n(x,y) = \mathrm{ORT}_{64n}(x^{64}, y^{64})$, that is:

$$f_n(x,y) = \begin{cases} -1 & \text{if } |\langle x,y\rangle| \leq \sqrt{n}/8 \\ +1 & \text{if } |\langle x,y\rangle| \geq \sqrt{n}/4. \end{cases}$$

In [She12], Sherstov proves the following result.

Lemma 5.5.12 ([She12])**.** *Let $\delta > 0$ be a sufficiently small constant and μ the uniform measure over $\{0,1\}^n \times \{0,1\}^n$. Then, $\mu(f_n^{-1}(+1)) = \Theta(1)$ and for all rectangle R in $\{0,1\}^n \times \{0,1\}^n$ such that $\mu(R) > 2^{-\delta n}$,*

$$\mu(R \cap f_n^{-1}(+1)) \geq \delta\mu(R \cap f_n^{-1}(-1)).$$

This implies that if we put uniform weight on inputs of ORT_{64n} of the form (x^{64}, y^{64}) and put 0 weight on the others, we get a distribution μ' satisfying the constraints of Theorem 5.5.3 for ORT_{64n} together with $\gamma = \delta$ from Lemma 4 and $g(64n) = 2^{\delta n}$.

To get a distribution satisfying the constraints of Theorem 5.5.3 on inputs of ORT_{64n+l} for all $0 \leq l \leq 63$ we extend μ' as follows:

$$\tilde{\mu}(xu, yv) = \begin{cases} \mu'(x,y) & \text{if } u = +1^l \text{ and } v = -1^l \text{ and } \left(\langle x,y\rangle < -\sqrt{64n} \text{ or } 0 \leq \langle x,y\rangle \leq \sqrt{64n}\right) \\ \mu'(x,y) & \text{if } u = +1^l \text{ and } v = +1^l \text{ and } \left(-\sqrt{64n} \leq \langle x,y\rangle < 0 \text{ or } \langle x,y\rangle > \sqrt{64n}\right) \\ 0 & \text{otherwise} \end{cases}$$

Using this distribution $\tilde{\mu}$ together with $\gamma = \delta$ from Lemma 5.5.12 and with $g(n) = 2^{-\delta n}$ we obtain, from Theorem 5.5.3, a Bell inequality violation for ORT_{64n+l} for all $0 \leq l \leq 63$:

Corollary 5.5.13. *There exists a Bell inequality B satisfying: $\forall \ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$,*

$$B(\mathbf{p}_{\mathrm{ORT}_{64n+l}}) = 2^{\delta n}\delta\tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(-1)),$$

and for any distribution $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}_{\mathrm{ORT}_{64n+l}}|_1 \leq \varepsilon$,

$$B(\mathbf{p}') \geq 2^{\delta n}\left(\delta\tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(-1)) - \epsilon\left[\delta\tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(-1)) + \tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(+1))\right]\right).$$

More precisely, Theorem 5.5.3 gives an explicit construction of such a Bell inequality: we can define B as:

$$B_{a,b,x,y} = \begin{cases} -2^{\delta n}\tilde{\mu}(x,y) & \text{if } (x,y) \in \mathrm{ORT}_{64n+l}^{-1}(+1) \text{ and } a \oplus b = -1 \\ \delta 2^{\delta n}\tilde{\mu}(x,y) & \text{if } (x,y) \in \mathrm{ORT}_{64n+l}^{-1}(-1) \text{ and } a \oplus b = -1 \\ 0 & \text{otherwise.} \end{cases}$$

Combining Corollary 5.5.13 together with the fact that $Q_{\varepsilon'}(\mathrm{ORT}_n) = O(\sqrt{n}\log n)$ [BCW98] we obtain, through Corollary 5.3.3, the following:

Proposition 5.5.14. *There is a quantum distribution $\overline{\mathbf{q}}_{\mathrm{ORT}} \in \mathcal{Q}$ and an explicit Bell inequality B satisfying: $B(\overline{\mathbf{q}}_{\mathrm{ORT}}) = 2^{\Omega(n)-O(\sqrt{n}\log n)}$, and for all $\ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$.*

5.6 Discussion

We have given three main results. First, we showed that normalized Bell inequalities can be modified to be bounded in absolute value on the larger set of local distributions that can abort without significantly changing the value of the violations achievable with non-signaling distributions. Then, we showed how to derive large inefficiency-resistant Bell violations from any gap between the partition bound and the quantum communication complexity of some given distribution $\mathbf{p}$. The distributions $\mathbf{q}$ achieving the large violations are relatively simple (only 3 outputs for boolean distributions $\mathbf{p}$) and can be made resistant to uniform noise at the expense of an increase in the number of outputs exponential in $Q(\mathbf{p})$. Finally, we showed how to construct explicit Bell inequalities when the separation between classical and quantum communication complexity is proven via the corruption bound.

From a practical standpoint, the specific Bell violations we have studied are probably not feasible to implement, because the parameters needed are still impractical or the quantum states are infeasible to implement. However, our results suggest that we could turn our attention to functions with small gaps in communication complexity, in order to find practical Bell inequalities that are robust against uniform noise and detector inefficiency.

To be more specific, let us consider an experimental setup with non-abort probability η per side, and ε uniform noise. Suppose we have a Boolean function with a lower bound of $c > 3\log(1/\eta^2)$ on classical communication with ε' error, and an $(\varepsilon'-\varepsilon)$-correct quantum protocol, with $\varepsilon' > \varepsilon$, using $q = \log(1/\eta^2)$ qubits. Our construction gives an inefficiency-resistant Bell violation of $2^{c-2q} > 1/\eta^2$, which is robust against ε uniform noise (the number of outcomes per side increases to $\frac{2}{\eta^2}$). Factoring in the inefficiency, the observed violation would still be $\eta^2 2^{c-2q} > 1$.

RESUMEN DEL CAPÍTULO

La cuestión de alcanzar grandes violaciones de Bell ha sido estudiada desde el articulo seminal de Bell en 1964 [Bel64]. En una linea de investigación, propuestas han sido hechas para exhibir familias de distribuciones que admiten violaciones no acotadas [Mer90, LPŻB04, NLP06, PGWP$^+$08]. En otra, varias medidas de no-localidad han sido estudiadas, como ser la cantidad de comunicación necesaria y suficiente para simular distribuciones cuánticas clásicamente[Mau92, BCT99, Ste00, TB03a, Pir03, DKLR11], o la resistencia a ineficiencias de detección y ruido. Más recientemente, el foco ha pasado a dar cotas superiores e inferiores en las violaciones alcanzables, en términos de varios parámetros: cantidad de jugadores, cantidad de entradas, cantidad de salidas, dimensión del estado cuántico, y nivel de entrelazamiento [DKLR11, JPPG$^+$10b, JP11].

Hasta bastante recientemente, las violaciones fueron estudiadas en el caso de distribuciones específicas (midiendo estados de Bell), o familias de distribuciones. Buhrman *et al.* [BRSdW12] dio una construcción que puede ser aplicada a varios problemas que tienen protocolos de comunicación cuántica eficientes (Definition 1.10.7), y para los cuales uno puede mostrar un trade-off entre comunicación y error en el setting clásico. Esto todavía requería un análisis *ad hoc* de problemas de comunicación. Recientemente Buhrman *et al.* [BCG$^+$16] propusieron la primera construcción general de estados cuánticos junto con desigualdades de Bell a partir de cualquier problema de comunicación. Los estados cuánticos violan las desigualdades de Bell cuando hay un espacio suficientemente grande entre la complejidad comunicacional cuántica y clásica (se necesita un espacio supra-cuadrático, al menos que exista un protocolo cuántico sin memoria local).

Volvemos al asunto de alcanzar grandes violaciones de Bell al explotar conexiones conocidas con complejidad comunicacional. Cotas inferiores fuertes en complejidad comunicacional, equivalentes a la cota de partición [JK10], equivalen a encontrar *desigualdades de Bell resistentes a ineficiencias* [LLR12]. Estas son funciones de Bell que están acotadas superiormente por 1 en todas las distribuciones locales *que pueden abortar*.

Primero estudiamos la resistencia de desigualdades de Bell a la ineficiencia. Mostramos que, hasta un factor constante en el valor de la violación, cualquier desigualdad de Bell normalizada puede ser hecha resistente a ineficiencia a la vez que mantiene la propiedad de normalización (**Theorem 5.2.1**).

Segundo, mostramos cómo derivar grandes violaciones de Bell a partir de cualquier problema de comunicación para el cual la cota de partición está acotada inferiormente y la complejidad computacional cuántica está acotada superiormente. Los problemas estudiados en complejidad computacional se encuentran mucho más allá del conjunto cuántico, pero mostramos cómo derivar fácilmente una distribución cuántica a partir de un protocolo cuántico. El valor de Bell que obtenemos es 2^{c-2q}, donde c es la cota inferior de partición en la complejidad comunicacional clásica del problema considerado, y q es una cota superior sobre su complejidad de comunicación cuántica (**Theorem 5.3.2 y Corollary 5.3.3**). La distribución cuántica tiene una salida extra por cada jugador comparada con la distribución original y usa la misma

cantidad de entrelazamiento que el protocolo cuántico y tantos pares EPR como son necesarios para teleportar en el protocolo la comunicación cuántica. Mostramos que estas violaciones de Bell pueden ser hechas resistentes al ruido, al costo de un factor 2^{2q} en el número de salidas por jugador (**Theorem 5.4.1**).

Finalmente, proveemos herramientas para construir desigualdades de Bell a partir de cotas inferiores de comunicación en la literatura. Cotas inferiores usadas en la práctica para separar complejidad comunicacional clásica de la cuántica usualmente son conseguidas usando cotas de corrupción y sus variantes. En **Theorem 5.5.3**, damos una construcción explícita que traduce estas cotas en un adecuado funcional de Bell. La tabla 5.3 resume los nuevos resultados o mejoras que obtenemos en este trabajo.

Problema	Violaciones normalizadas de Bell [BCG$^+$16]	Violaciones de Bell resistentes a ineficiencia (este trabajo)
VSP [Raz99a, KR11]	$\Omega\left(\sqrt[6]{n}/\sqrt{\log n}\right)$ $d = 2^{\Theta(n \log n)}, K = 2^{\Theta(n)}$	$2^{\Omega(\sqrt[3]{n})-O(\log n)}$ $d = 2^{O(\log n)}, K = 3$
DISJ [Raz92, Raz03, AA05]	N/A	$2^{\Omega(n)-O(\sqrt{n})}$ $d = 2^{O(\sqrt{n})}, K = 3$
TRIBES [JKS03, BCW98]	N/A	$2^{\Omega(n)-O(\sqrt{n}\log^2 n)}$ $d = 2^{O(\sqrt{n}\log^2 n)}, K = 3$
ORT [She12, BCW98]	N/A	$2^{\Omega(n)-O(\sqrt{n}\log n)}$ $d = 2^{O(\sqrt{n}\log n)}, K = 3$

Tab. 5.3: Comparación de las violaciones de Bell obtenidas por la construcción general de Buhrman *et al.* [BCG$^+$16] para violaciones de Bell normalizadas (segunda columna) y este trabajo, para violaciones de Bell resistentes a ineficiencia (ver Propositions 5.5.4, 5.5.8, 5.5.11, and 5.5.14), en términos de la dimensión d del espacio de Hilbert local, el tamaño n de conjuntos de configuración de mediciones (o entradas) (típicamente $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$) y el número de resultados K (o salidas) por parte de las distribuciones cuánticas. Desigualdades de Bell explícitas son dadas en la Sección5.5.2. La construcción de Buhrman *et al.* solo da una violación cuando el espacio entre las complejidades clásica y cuánticas es mayor que cuadrática. En el caso en que el espacio sea demasiado pequeño como para provar una violación, indicamos esto con "N/A".

6. OPEN QUESTIONS AND FUTURE RESEARCH

We close this thesis with a list of the main questions that remain open in each chapter.

In Chapter 2, we have shown that if at least one of the players in a bipartite Bell experiment uses a computable function f to choose his inputs, then an eavesdropper without knowledge of f can prepare seemingly non-local boxes provided she knows a computable upper bound on the time computational complexity of f. In spite of the fact that every computable function f is computable $O(T(n))$-time for some computable time bound T (furthermore, as discussed before, such a bound can be derived from physical assumptions on the computing devices) and therefore every Bell experiment with computable inputs is subject to our loophole, it is natural to wonder if we can give a strategy for the eavesdropper which is independent of the running time of the target f.

Second, in Chapter 3, we have shown that any model of Nature predicting non-local correlations between the outputs of distributed computing devices, linked via some kind of instantaneous hidden-signaling mechanism, is in conflict with special relativity. As in Chapter 2, we have an assumption about the time computational complexity of the devices, this time to prove the soundness of the protocol derived. Therefore, again one can study the possibility of getting rid of that assumption. Furthermore, it is interesting to study if even with non-computable devices one can still have signaling.

Third, in Chapter 4, we have shown that no amount of pseudorandomness is enough to prepare a maximally mixed state as a mixture of pure states. The distinguishing procedures given, although sufficient to reach the theoretical result, are far from being efficient and therefore, from a cryptographic perspective, unfair (in cryptography, Bob, being the adversary, is usually limited to polynomial resources). Hence, a future line of research could be to come up with *efficient* distinguishing protocols.

Finally, in Chapter 5, we have shown how to derive quantum Bell violations, resistant to the detection loophole and uniform noise, from any gap between the partition bound and quantum communication complexity (QCC). Although it is one the largest lower bounds on classical communication complexity (CC) so far, we know that, for low complexities, the partition bound is not tight [GJPW15]. Hence, it is possible that the general construction of [BCG$^+$16] (requiring, as the reader may recall, a quadratic gap between CC and QCC) applies to functions for which ours does not. Therefore, it is interesting to investigate whether we can come up with a general construction, via CC lower bounds, that works for every gap between CC and QCC. Or, at least, one that works when there is a, say, quadratic gap but has the nice properties in terms of resource usage (number of outputs, amount of entanglement, etc) that ours have. Another open question relates to the study of upper bounds for the quantum violations of Bell inequalities. For normalized Bell inequalities, it is known that the maximum violation is upper bounded by $\max\{d, K, N\}$ with d the local dimension of the Hilbert spaces and N (resp. K) the number of inputs (resp. outputs) per side (see Table 5.2). For the quantum violations of inefficiency-resistant Bell inequalities, we have an upper bound of N

and we know that they are not bounded by the K (we have given exponential violations with $K = 3$) but we do not know any bound in terms of d.

RESUMEN DEL CAPÍTULO

Cerramos esta tesis con una lista de las principales preguntas abiertas que surgen de cada capítulo.

Capítulo 2. En el capítulo 2, hemos mostrado que si al menos uno de los jugadores en un experimento bipartito de Bell usa una función computable f para elegir sus entradas, entonces un espía sin conocimiento de f puede preparar cajas aparentemente no-locales siempre y cuando sepa una cota superior computable a la complejidad temporal de f.

A pesar de que toda función computable f es computable en tiempo $O(T(n))$ para alguna cota temporal computable T (más aún, como se discutió, tal cota puede ser derivada de hipotesis físicas sobre los dispositivos computacionales intervinientes) y, por lo tanto, todo experimento de Bell con entradas computables es pasible de nuestro loophole, es natural preguntarse si se puede dar una estrategia para el espía que sea independiente de la complejidad temporal de f.

Capítulo 3. En el capítulo 3, hemos mostrado que cualquier modelo para la Naturaleza que prediga correlaciones no-locales entre las salidas de dispositivos computacionales distribuidos, vinculados a través de algún tipo de mecanismo de señalización instantáneo, está en conflicto con la relatividad especial.

Así como en el capítulo 2, tenemos una hipótesis sobre la complejidad temporal de los dispositivos, esta vez para probar la correctitud del protocolo presentado. Por lo tanto, de nuevo uno puede preguntarse acerca de la posibilidad de deshacerse de tal hipótesis. Más aún, sería interesante estudiar si aún con dispositivos no-computables uno todavía puede tener señalización.

Capítulo 4. En el capítulo 4, hemos mostrado que no hay cantidad de pseudoaleatoriedad suficiente que permita preparar el estado máximamente mixto como una mezcla de estados puros.

El protocolo de realizar la distinción presentado, aunque suficiente para concluir el resultado teórico, está lejos de ser eficiente y entonces, desde una perspectiva criptográfica, es injusto (en criptografía, Bob, el adversario, está usualmente limitado a recursos polinomiales). Por lo tanto, una linea de investigación futura podría ser la de tratar de encontrar protocolos de distinción eficientes.

Capítulo 5. Finalmente, en el capítulo 5, hemos mostrado cómo derivar violaciones cuánticas de Bell, resistentes al loophole de la detección y a ruido uniforme, de cualquier separación entre la partition-bound y la complejidad comunicacional cuántica (CCQ).

A pesar de que es de las cotas inferiores a la complejidad comunicacional clásica (CC) más ajustadas que se han desarrollado hasta el momento, sabemos que, para complejidades

bajas, la partition-bound coincide con la estrictamente menor a la CC [GJPW15]. Por lo tanto, es posible que la construcción general de Buhrman et al. [BCG$^+$16] (la cual requiere, como el lector recordará, una separación cuadrática entre QCC y CCC), aplique a funciones para las cuales nuestra construcción no aplica. En consecuencia, sería interesante investigar la posibilidad de encontrar una construcción general, via cotas inferiores a la CC, que funcione para cualquier gap entre QCC y CC; o, por lo menos, una que funcione para cuando hay, digamos, una separación cuadrática pero que tengas las buenas propiedades en términos de utilización de recursos (número de salidas, cantidad entrelazamiento, etc) que tiene nuestra construcción.

Otra pregunta abierta tiene que ver con el estudio de cotas superiores a las violaciones cuánticas de desigualdades de Bell. Para desigualdades de Bell normalizadas, se sabe que la máxima violación está acotada superiormente por $\max\{d, K, N\}$, con d la dimensión de los espacios de Hilbert locales y N (resp. K) la cantidad de entradas (resp. salidas) por lado (ver la Tabla 5.2). Para la violación máxima de desigualdades de Bell resistentes-a-ineficiencias, tenemos una cota superior de N y sabemos que no están acotadas por K (dimos violaciones exponenciales en $QCC - CC$ con $K = 3$), pero no sabemos si están o no acotadas por alguna función de d.

BIBLIOGRAPHY

[AA05] S. Aaronson and A. Ambainis, *Quantum search of spatial regions*, Theory of Computing **1** (2005), 47–79.

[AAM+15] Carlos Abellán, Waldimar Amaya, Daniel Mitrani, Valerio Pruneri, and Morgan W Mitchell, *Generation of fresh and pure random numbers for loophole-free bell tests*, Physical Review Letters **115** (2015), no. 25, 250403.

[AB91] Leonard M Adleman and Manuel Blum, *Inductive inference and unsolvability*, The Journal of Symbolic Logic **56** (1991), no. 03, 891–900.

[AB09a] Elias Amselem and Mohamed Bourennane, *Experimental four-qubit bound entanglement*, Nature Physics **5** (2009), no. 10, 748–752.

[AB09b] Sanjeev Arora and Boaz Barak, *Computational complexity: a modern approach*, Cambridge University Press, 2009.

[ACCS12] Alastair A Abbott, Cristian S Calude, Jonathan Conder, and Karl Svozil, *Strong kochen-specker theorem and incomputability of quantum randomness*, Physical Review A **86** (2012), no. 6, 062109.

[ACS15] Alastair A. Abbott, Cristian S. Calude, and Karl Svozil, *A variant of the Kochen-Specker theorem localising value indefiniteness*, Journal of Mathematical Physics **56** (2015), no. 10, 102201(1–17).

[ADGL02] A. Acín, T. Durt, N. Gisin, and J. I. Latorre, *Quantum nonlocality in two three-level systems*, Physical Review A **65** (2002), 052325.

[ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger, *Experimental test of bell's inequalities using time-varying analyzers*, Physical Review Letters **49** (1982), no. 25, 1804.

[BB84] C. H. Bennett and G. Brassard, *Quantum cryptography: Public key distribution and coin tossing*, Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India), 1984, p. 175.

[BCG+16] H. Buhrman, Ł. Czekaj, A. Grudka, Mi. Horodecki, P. Horodecki, M. Markiewicz, F. Speelman, and S. Strelchuk, *Quantum communication complexity advantage implies violation of a Bell inequality*, Proceedings of the National Academy of Sciences **113** (2016), no. 12, 3191–3196.

[BCH+02] Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu, *Quantum nonlocality, bell inequalities, and the memory loophole*, Physical Review A **66** (2002), no. 4, 042111.

[BCP+14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner, *Bell nonlocality*, Reviews of Modern Physics **86** (2014), no. 2, 419.

[BCS+04] MD Barrett, J Chiaverini, T Schaetz, J Britton, WM Itano, JD Jost, E Knill, C Langer, D Leibfried, R Ozeri, et al., *Deterministic quantum teleportation of atomic qubits*, Nature **429** (2004), no. 6993, 737–739.

[BCSS11] N. Brunner, D. Cavalcanti, A. Salles, and P. Skrzypczyk, *Bound nonlocality and activation*, Physical Review Letters **106** (2011), 020402.

[BCT99] G. Brassard, R. Cleve, and A. Tapp, *Cost of exactly simulating quantum entanglement with classical communication*, Physical Review Letters **83** (1999), no. 9, 1874.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson, *Quantum vs classical communication and computation*, Proceedings of the 30th Annual ACM Symposium on Theory of Computing, 1998, pp. 63–68.

[BdlTS+16] Ariel Bendersky, Gonzalo de la Torre, Gabriel Senno, Santiago Figueira, and Antonio Acín, *Algorithmic pseudorandomness in quantum setups*, Physical Review Letters **116** (2016), 230402.

[BdlTS+17] Ariel Bendersky, Gonzalo de la Torre, Gabriel Senno, Santiago Figueira, and Antonio Acin, *Non-signaling deterministic models for non-local correlations have to be uncomputable*, Physical Review Letters (2017), to appear.

[Bel64] John S Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics **1** (1964), no. 3, 195–200.

[BFS86] L. Babai, P. Frankl, and J. Simon, *Complexity classes in communication complexity theory*, Proc. 27th FOCS, IEEE, 1986, pp. 337–347.

[BG11] Jonathan Barrett and Nicolas Gisin, *How much measurement independence is needed to demonstrate nonlocality?*, Physical Review Letters **106** (2011), no. 10, 100406.

[Boh52] David Bohm, *A suggested interpretation of the quantum theory in terms of" hidden" variables. i*, Physical Review **85** (1952), no. 2, 166.

[BPSW06] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson, *A strong direct product theorem for corruption and the multiparty communication complexity of disjointness*, Computational Complexity **15** (2006), no. 4, 391–432.

[BRSdW12] H. Buhrman, O. Regev, G. Scarpa, and R. de Wolf, *Near-optimal and explicit Bell inequality violations*, Theory of Computing **8** (2012), no. 1, 623–645.

[BSHC85] John S Bell, Abner Shimony, Michael A Horne, and John F Clauser, *An exchange on local beables*, Dialectica **39** (1985), no. 2, 85–110.

[BY08] Adam Brandenburger and Noson Yanofsky, *A classification of hidden-variable properties*, Journal of Physics A: Mathematical and Theoretical **41** (2008), no. 42, 425302.

[BYJK04] Ziv Bar-Yossef, Thathachar S Jayram, and Iordanis Kerenidis, *Exponential separation of quantum and classical one-way communication complexity*, Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, ACM, 2004, pp. 128–137.

[CH74] John F. Clauser and Michael A. Horne, *Experimental consequences of objective local theories*, Physical Review D **10** (1974), 526–535.

[Cha75] Gregory J Chaitin, *A theory of program size formally identical to information theory*, Journal of the ACM (JACM) **22** (1975), no. 3, 329–340.

[CHSH69] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt, *Proposed experiment to test local hidden-variable theories*, Physical Review Letters **23** (1969), no. 15, 880.

[Cir80] Boris S Cirel'son, *Quantum generalizations of Bell's inequality*, Letters in Mathematical Physics **4** (1980), no. 2, 93–100.

[CK11] Roger Colbeck and Adrian Kent, *Private randomness expansion with untrusted devices*, Journal of Physics A: Mathematical and Theoretical **44** (2011), no. 9, 095305–.

[CR12] Roger Colbeck and Renato Renner, *Free randomness can be amplified*, Nature Physics **8** (2012), no. 6, 450–454.

[CvDNT99] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp, *Quantum entanglement and the communication complexity of the inner product function*, Quantum Computing and Quantum Communications, Springer Berlin Heidelberg, 1999, pp. 61–74.

[DKLR11] J. Degorre, M. Kaplan, S. Laplante, and J. Roland, *The communication complexity of non-signaling distributions*, Quantum information & computation **11** (2011), no. 7-8, 649–676.

[Ebe93] Philippe H Eberhard, *Background level and counter efficiencies required for a loophole-free einstein-podolsky-rosen experiment*, Physical Review A **47** (1993), no. 2, R747.

[Eke91] A. K. Ekert, *Quantum cryptography based on bell's theorem*, Physical Review Letters **67** (1991), 661.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Physical review **47** (1935), no. 10, 777.

[Fin82] Arthur Fine, *Hidden variables, joint probability, and the Bell inequalities*, Physical Review Letters **48** (1982), no. 5, 291.

[FN15] Santiago Figueira and André Nies, *Feasible analysis, randomness, and base invariance*, Theory of Computing Systems **56** (2015), no. 3, 439–464.

[FWW09] M. Forster, S. Winkler, and S. Wolf, *Distilling nonlocality*, Physical Review Letters **102** (2009), 120401.

[GJPW15] Mika Göös, TS Jayram, Toniann Pitassi, and Thomas Watson, *Randomized communication vs. partition number.*, 2015.

[GMDLT+13] Rodrigo Gallego, Lluis Masanes, Gonzalo De La Torre, Chirag Dhara, Leandro Aolita, and Antonio Acín, *Full randomness from arbitrarily deterministic events*, Nature communications **4** (2013).

[Gol67] E Mark Gold, *Language identification in the limit*, Information and control **10** (1967), no. 5, 447–474.

[Gro53] A. Grothendieck, *Résumé de la théorie métrique des produits tensoriels topologiques*, Boletim da Sociedade de Matemática de São Paulo **8** (1953), 1–79.

[GVW+15] Marissa Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger, *Significant-loophole-free test of Bell's theorem with entangled photons*, Physical Review Letters **115** (2015), 250401.

[GWAN12] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, *Operational framework for nonlocality*, Physical Review Letters **109** (2012), 070401.

[Hal10] Michael JW Hall, *Local deterministic model of singlet state correlations based on relaxing measurement independence*, Physical review letters **105** (2010), no. 25, 250404.

[HBD+15] Bas Hensen, H Bernien, AE Dréau, A Reiserer, N Kalb, MS Blok, J Ruitenberg, RFL Vermeulen, RN Schouten, and C ABellán, *Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres*, Nature **526** (2015), no. 7575, 682–686.

[HJ13] P. Harsha and R. Jain, *A Strong Direct Product Theorem for the Tribes Function via the Smooth-Rectangle Bound*, Proceedings of the 33rd International Conference on Foundations of Software Technology and Theoretical Computer Science, vol. 24, 2013, pp. 141–152.

[Hol73] Alexander Semenovich Holevo, *Bounds for the quantity of information transmitted by a quantum communication channel*, Problemy Peredachi Informatsii **9** (1973), no. 3, 3–11.

[IH08] Satoshi Ishizaka and Tohya Hiroshima, *Asymptotic teleportation scheme as a universal programmable quantum processor*, Physical Review Letters **101** (2008), no. 24, 240501.

[IH09] ———, *Quantum teleportation scheme by selecting one of multiple output ports*, Physical Review A **79** (2009), no. 4, 042306.

[Jar84] Jon P Jarrett, *On the physical significance of the locality conditions in the bell arguments*, Noûs (1984), 569–589.

[JK10] R. Jain and H. Klauck, *The partition bound for classical communication complexity and query complexity*, Proceedings of the 25th IEEE Annual Conference on Computational Complexity, 2010, pp. 247–258.

[JKS03] T. S. Jayram, R. Kumar, and D. Sivakumar, *Two applications of information complexity*, Proceedings of the 35th Annual ACM Symposium on Theory of Computing, 2003, pp. 673–682.

[JP11] M. Junge and C. Palazuelos, *Large violation of Bell inequalities with low entanglement*, Communications in Mathematical Physics **306** (2011), no. 3, 695–746.

[JPPG+10a] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf, *Operator space theory: A natural framework for Bell inequalities*, Physical Review Letters **104** (2010), 170405.

[JPPG+10b] ———, *Unbounded violations of bipartite Bell inequalities via operator space theory*, Communications in Mathematical Physics **300** (2010), no. 3, 715–739.

[KGZ+00] D. Kaszlikowski, P. Gnaciński, M. Żukowski, W. Miklaszewski, and A. Zeilinger, *Violations of local realism by two entangled N-dimensional systems are stronger than for two qubits*, Physical Review Letters **85** (2000), 4418–4421.

[Kle38] Stephen Cole Kleene, *On notation for ordinal numbers*, The Journal of Symbolic Logic **3** (1938), no. 04, 150–155.

[KLL+15] I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao, *Lower bounds on information complexity via zero-communication protocols and applications*, SIAM Journal on Computing **44** (2015), no. 5, 1550–1572.

[KMSY14] G. Kol, S. Moran, A. Shpilka, and A. Yehudayoff, *Approximate nonnegative rank is equivalent to the smooth rectangle bound*, Automata, Languages, and Programming, Springer, 2014, pp. 701–712.

[KN97] E. Kushilevitz and N. Nisan, *Communication complexity*, Cambridge University Press, 1997.

[KR11] B. Klartag and O. Regev, *Quantum one-way communication can be exponentially stronger than classical communication*, Proceedings of the 43th Annual ACM Symposium on Theory of Computing, 2011, pp. 31–40.

[LG04] J-Å. Larsson and R. D. Gill, *Bell's inequality and the coincidence-time loophole*, EPL (Europhysics Letters) **67** (2004), no. 5, 707.

[LGSdlT+] Ignacio H. López Grande, Gabriel Senno, Gonzalo de la Torre, Miguel A. Larotonda, Ariel Bendersky, Santiago Figueira, and Antonio Acín, *Distinguishing computable mixtures of quantum states*, Submitted.

[LGSL16] Ignacio H. López Grande, Christian T. Schmiegelow, and Miguel A. Larotonda, *Autonomous open-source hardware apparatus for quantum key distribution*, Papers in Physics **8** (2016), 080002.

[LKPR10] Jonathan Lavoie, Rainer Kaltenbaek, Marco Piani, and Kevin J Resch, *Experimental bound entanglement in a four-photon state*, Physical Review Letters **105** (2010), no. 13, 130501.

[LLN+16] Sophie Laplante, Mathieu Laurière, Alexandre Nolin, Jérémie Roland, and Gabriel Senno, *Robust Bell Inequalities from Communication Complexity*, 11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016) (Dagstuhl, Germany) (Anne Broadbent, ed.), Leibniz International Proceedings in Informatics (LIPIcs), vol. 61, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016, pp. 5:1–5:24.

[Llo00] Seth Lloyd, *Ultimate physical limits to computation*, Nature **406** (2000), no. 6799, 1047–1054.

[LLR12] S. Laplante, V. Lerays, and J. Roland, *Classical and quantum partition bound and detector inefficiency*, Proceedings of the 39th International Colloquium on Automata, Languages and Programming, 2012, pp. 617–628.

[LPŻB04] W. Laskowski, T. Paterek, M. Żukowski, and Č Brukner, *Tight multipartite Bell's inequalities involving many measurement settings*, Physical Review Letters **93** (2004), no. 20, 200401.

[LS09] N. Linial and A. Shraibman, *Lower bounds in communication complexity based on factorization norms*, Random Structures & Algorithms **34** (2009), no. 3, 368–394.

[Mas02] S. Massar, *Nonlocality, closing the detection loophole, and communication complexity*, Physical Review A **65** (2002), 032121.

[Mau92] T. Maudlin, *Bell's inequality, information transmission, and prism models*, PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, JSTOR, 1992, pp. 404–417.

[Mer90] D. N. Mermin, *Extreme quantum entanglement in a superposition of macroscopically distinct states*, Physical Review Letters **65** (1990), no. 15, 1838.

[ML66] Per Martin-Löf, *The definition of random sequences*, Information and control **9** (1966), no. 6, 602–619.

[MN98] Makoto Matsumoto and Takuji Nishimura, *Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator*, ACM Transactions on Modeling and Computer Simulation (TOMACS) **8** (1998), no. 1, 3–30.

[MP03a] Serge Massar and Stefano Pironio, *Violation of local realism versus detection efficiency*, Physical Review A **68** (2003), 062109.

[MP03b] ———, *Violation of local realism versus detection efficiency*, Physical Review A **68** (2003), no. 6, 062109.

[MPA11] Lluís Masanes, Stefano Pironio, and Antonio Acín, *Secure device-independent quantum key distribution with causally independent measurement devices*, Nature communications **2** (2011), 238.

[MPRG02] S. Massar, S. Pironio, J. Roland, and B. Gisin, *Bell inequalities resistant to detector inefficiency*, Physical Review A **66** (2002), 052112.

[NC11] M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information: 10th anniversary edition*, Quantum Computation and Quantum Information, Cambridge University Press, 2011.

[Nie09] André Nies, *Computability and randomness*, vol. 51, Oxford University Press, 2009.

[NLP06] K. Nagata, W. Laskowski, and T. Paterek, *Bell inequality with an arbitrary number of settings and its applications*, Physical Review A **74** (2006), no. 6, 062109.

[Nor11] Travis Norsen, *John S. Bell's concept of local causality*, American Journal of Physics **79** (2011), no. 12, 1261–1275.

[Odi92] Piergiorgio Odifreddi, *Classical recursion theory: The theory of functions and sets of natural numbers*, Elsevier, 1992.

[PAM+10] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzimitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al., *Random numbers certified by bell's theorem*, Nature **464** (2010), no. 7291, 1021–1024.

[PGWP+08] D. Pérez-García, M. M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge, *Unbounded violation of tripartite Bell inequalities*, Communications in Mathematical Physics **279** (2008), no. 2, 455–486.

[Pir03] S. Pironio, *Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations*, Physical Review A **68** (2003), no. 6, 062102.

[PM13] Stefano Pironio and Serge Massar, *Security of practical private randomness generation*, Physical Review A **87** (2013), no. 1, 012336.

[PR94] Sandu Popescu and Daniel Rohrlich, *Quantum nonlocality as an axiom*, Foundations of Physics **24** (1994), no. 3, 379–385.

[PRB+14] Gilles Pütz, Denis Rosset, Tomer Jack Barnea, Yeong-Cherng Liang, and Nicolas Gisin, *Arbitrarily small amount of measurement independence is sufficient to manifest quantum nonlocality*, Physical Review Letters **113** (2014), 190402.

[PWT+07] Robert Prevedel, Philip Walther, Felix Tiefenbacher, Pascal Böhi, Rainer Kaltenbaek, Thomas Jennewein, and Anton Zeilinger, *High-speed linear optics quantum computing using active feed-forward*, Nature **445** (2007), no. 7123, 65–69.

[Raz92] A. A. Razborov, *On the distributional complexity of disjointness*, Theoretical Computer Science **106** (1992), no. 2, 385 – 390.

[Raz99a] R. Raz, *Exponential separation of quantum and classical communication complexity*, Proceedings of the 31st Annual ACM Symposium on Theory of Computing, 1999, pp. 358–367.

[Raz99b] Ran Raz, *Exponential separation of quantum and classical communication complexity*, Proceedings of the thirty-first annual ACM symposium on Theory of computing, ACM, 1999, pp. 358–367.

[Raz03] A. A. Razborov, *Quantum communication complexity of symmetric predicates*, Izvestiya: Mathematics **67** (2003), no. 1, 145.

[RKM+01] Mary A Rowe, David Kielpinski, Volker Meyer, Charles A Sackett, Wayne M Itano, Christopher Monroe, and David J Wineland, *Experimental violation of a bell's inequality with efficient detection*, Nature **409** (2001), no. 6822, 791–794.

[RT09] Oded Regev and Ben Toner, *Simulating quantum correlations with finite communication*, SIAM Journal on Computing **39** (2009), no. 4, 1562–1580.

[Sch71] Claus-Peter Schnorr, *Zufälligkeit und Wahrscheinlichkeit*, vol. 218, Springer-Verlag, Heidelberg, 1971.

[She12] A. A. Sherstov, *The communication complexity of gap hamming distance.*, Theory of Computing **8** (2012), no. 1, 197–208.

[Shi86] Abner Shimony, *Events and processes in the quantum world*, Quantum Concepts in Space and Time (Roger Penrose and C. J. Isham, eds.), New York; Oxford University Press, 1986, pp. 182–203.

[Sho97] P.W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM J. Comput. **26** (1997), 1484–1509.

[SMSC+15] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, Scott Glancy, Deny R. Hamel, Michael S. Allman, Kevin J. Coakley, Shellee D. Dyer, Carson Hodge, Adriana E. Lita, Varun B. Verma, Camilla Lambrocco, Edward Tortorici, Alan L. Migdall, Yanbao Zhang, Daniel R. Kumor, William H. Farr, Francesco Marsili, Matthew D. Shaw, Jeffrey A. Stern, Carlos ABellán, Waldimar Amaya, Valerio Pruneri, Thomas Jennewein, Morgan W. Mitchell, Paul G. Kwiat, Joshua C. Bienfang, Richard P. Mirin, Emanuel Knill, and Sae Woo Nam, *Strong loophole-free test of local realism*, Physical Review Letters **115** (2015), 250402.

[Soa99] Robert I Soare, *Recursively enumerable sets and degrees: A study of computable functions and computably generated sets*, Springer Science & Business Media, 1999.

[Ste00] M. Steiner, *Towards quantifying non-local information transfer: finite-bit nonlocality*, Physics Letters A **270** (2000), no. 5, 239–244.

[SZ08] Yaoyun Shi and Yufan Zhu, *Tensor norms and the classical communication complexity of nonlocal quantum measurement*, SIAM Journal on Computing **38** (2008), no. 3, 753–766.

[TB03a] B. F. Toner and D. Bacon, *Communication cost of simulating Bell correlations*, Physical Review Letters **91** (2003), no. 18, 187904.

[TB03b] Benjamin F Toner and Dave Bacon, *Communication cost of simulating Bell correlations*, Physical Review Letters **91** (2003), no. 18, 187904.

[TBZG98] Wolfgang Tittel, Jürgen Brendel, Hugo Zbinden, and Nicolas Gisin, *Violation of bell inequalities by photons more than 10 km apart*, Physical Review Letters **81** (1998), no. 17, 3563.

[TDH+05] H Takesue, E Diamanti, T Honjo, C Langrock, MM Fejer, K Inoue, and Y Yamamoto, *Differential phase shift quantum key distribution experiment over 105 km fibre*, New Journal of Physics **7** (2005), no. 1, 232.

[TMF+13] Shuntaro Takeda, Takahiro Mizuta, Maria Fuwa, Peter van Loock, and Akira Furusawa, *Deterministic quantum teleportation of photonic quantum bits by a hybrid technique*, Nature **500** (2013), no. 7462, 315–318.

[TSS13] Le Phuc Thinh, Lana Sheridan, and Valerio Scarani, *Bell tests with min-entropy sources*, Physical Review A **87** (2013), 062121.

[Tur37] Alan Mathison Turing, *On computable numbers, with an application to the entscheidungsproblem*, Proceedings of the London mathematical society **2** (1937), no. 1, 230–265.

[Val02] Antony Valentini, *Signal-locality in hidden-variables theories*, Physics Letters A **297** (2002), 273–278.

[VPB10] Tamás Vértesi, Stefano Pironio, and Nicolas Brunner, *Closing the detection loophole in bell experiments using qudits*, Physical Review Letters **104** (2010), no. 6, 060401.

[Wie78] R Wiehagen, *Zur theorie der algorithmischen erkennung*, Ph.D. thesis, Humboldt-Universitat, Berlin, 1978.

[WJS+98] Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger, *Violation of bell's inequality under strict einstein locality conditions*, Physical Review Letters **81** (1998), no. 23, 5039.

[Wol15] Stefan Wolf, *Nonlocality without counterfactual reasoning*, Physical Review A **92** (2015), 052102.

[Yao79] Andrew Chi-Chih Yao, *Some complexity questions related to distributive computing (preliminary report)*, Proceedings of the eleventh annual ACM symposium on Theory of computing, ACM, 1979, pp. 209–213.

[Yao83] ———, *Lower bounds by probabilistic arguments*, Foundations of Computer Science, 1983., 24th Annual Symposium on, IEEE, 1983, pp. 420–428.

[Yao93] A. C. C. Yao, *Quantum circuit complexity*, Proceedings of the 34th Annual Symposium on Foundations of Computer Science, IEEE, 1993, pp. 352–361.

[Yur00] Ulvi Yurtsever, *Quantum mechanics and algorithmic randomness*, Complexity **6** (2000), no. 1, 27–34.

[ZZ08] Thomas Zeugmann and Sandra Zilles, *Learning recursive functions: A survey*, Theoretical Computer Science **397** (2008), no. 1, 4–56.