Quantifying randomness from Bell nonlocality

Boris Bourdoncle

7 January 2019

Abstract

The twentieth century was marked by two scientific revolutions. On the one hand, quantum mechanics questioned our understanding of nature and physics. On the other hand, came the realisation that information could be treated as a mathematical quantity. They together brought forward the age of information.

A conceptual leap took place in the 1980's, that consisted in treating information in a quantum way as well. The idea that the intuitive notion of information could be governed by the counter-intuitive laws of quantum mechanics proved extremely fruitful, both from a fundamental point of view, where an information-theoretic approach to quantum mechanics helps us shine a new light on quantum theory, and from a practical point of view, where the concept of quantum information gave birth to unforeseen ways of processing and transmitting information, such as quantum computing and quantum communication.

The notion of randomness plays a central role in that respect. Indeed, the laws of quantum physics are probabilistic: that contrasts with thousands of years of physical theories that aimed to derive deterministic laws of nature. This, in turn, provides us with sources of random numbers, a crucial resource for information protocols.

The fact that quantum theory only describes probabilistic behaviours was for some time regarded as a form of incompleteness: a more accurate description of the laws of quantum theory would make its predictions deterministic. But a specific property, Bell nonlocality, observed on some correlations predicted by quantum theory, showed that this approach was an impasse: the laws of quantum physics are inherently probabilistic, i.e., they cannot be completed in such a way that their apparent randomness could be traced back to a lack of knowledge.

This observation has practical consequences, as witnessing Bell nonlocality then certifies the presence of intrinsic randomness: the outputs of a nonlocal

physical process are necessarily unpredictable, and are, in that sense, truly random. Moreover, that assertion does not depend on the physical system from which nonlocality is observed. Certifying randomness from nonloality thus allows us to assess true randomness, that is, randomness that cannot be explained by ignorance, in a device-independent manner.

In this thesis, we quantify nonlocality-based randomness in various frameworks. In the first scenario, we quantify randomness without relying on the quantum formalism. We consider a nonlocal process and assume that it has a specific causal structure, that is only due to how it evolves with time. We provide trade-offs between nonlocality and randomness for the various causal structures that we consider.

Nonlocality-based randomness is usually defined in a theoretical framework: randomness is certified for an abstract, mathematical process. In the second scenario, we take a practical approach and ask how much randomness can be certified for a practical process, from which only partial knowledge can be gained with experiment. We describe a method to optimise how much randomness can be certified in such a situation.

Trade-offs between nonlocality and randomness are usually studied in the bipartite case, as two agents is the minimal requirement to define nonlocality. In the third scenario, we quantify how much randomness can be certified for a tripartite process. We also look into possible applications of such trade-offs.

Though nonlocality-based randomness is device-independent, the process from which randomness is certified is actually realised with a physical state. In the fourth scenario, we ask what physical requirements should be imposed on the physical state for maximal randomness to be certified, and more specifically, how entangled the underlying state should be. We show that maximal randomness can be certified from any level of entanglement.

Contents

List of publications

- Boris Bourdoncle, Stefano Pironio, and Antonio Acín. "Quantifying the randomness of copies of noisy Popescu-Rohrlich correlations". *Physical Review A* **98**, 042130, 2018.
- Boris Bourdoncle, Pei-Sheng Lin, Denis Rosset, Antonio Acín and Yeong-Cherng Liang. "Regularising data for practical randomness generation". arXiv:1802.04703, 2018.
- Erik Woodhead, Boris Bourdoncle, and Antonio Acín. "Randomness versus nonlocality in the Mermin-Bell experiment with three parties". *Quantum* **2**, 82, 2018.
- Erik Woohdead, Jedrzej Kaniewski, Boris Bourdoncle, Alexia Salavrakos, Joseph Bowles, Remigiusz Augusiak, and Antonio Acín. "Maximal randomness from partially entangled states". In preparation.

List of acronyms

CHSH	Clauser-Horn-Shimony-Holt
DI	Device-independent
i.i.d.	Independent and identically distributed
LP	Linear programming
NPA	Navascués-Pironio-Acín
POVM	Positive-operator valued measure
PR-box	Popescu-Rohrlich box
QKD	Quantum key distribution
RNG	Randomness generation
SDP	Semidefinite programming
SOS	Sum-of-square

Chapter 1

Introduction

We first give a short historical overview of the field of quantum information theory. We then describe what motivated us to study the questions presented in this Thesis and how we contributed to answering them.

1.1 The quantum and information revolutions

At the beginning of the twentieth century, the pioneering works of Max Planck and Albert Einstein were the beginning of a paradigm shift in the physics of microscopic scales: experimental observations, such as the black-body radiation or the photoelectric effect, were no longer compatible with classical mechanics. The new mathematical formalisms that were then developed by Werner Heisenberg, Erwin Schrödinger, Max Born and others founded what would become quantum mechanics. This new theory deeply modified our perception, not only of nature, but also of what a physical theory should be. Moreover, this novel understanding of the behaviour of particles at the microscopic level enabled the emergence of new technological devices such as lasers or semi-conductors.

In the 1930's and 40's, the groundbreaking ideas of Alan Turing and Claude Shannon led to the birth of computer science and information theory. Their conceptual works were concerned with how to process and communicate information. Based on this theoretical research, a myriad of information-related technological devices could then be developed. They were engineered thanks to electronic components whose functioning rely on quantum effects, such as transistors. This was the advent of the information age, that of cell phones, computers and the

Internet.

However, the information that can be encoded, processed and transmitted by such means is meant to be classical: a unit of information, the binary digit, or 'bit', takes the value '0' or '1'. It turns out that, if information is represented by a quantum bit, or 'qubit', the full potential of quantum physics translates into new possibilities for information tasks.

Indeed, a quantum bit, that is, a unit of information encoded on a quantum state, inherits the specificities of quantum physics, such as entanglement or state superposition. These revolutionary concepts, never contemplated until a century ago, are the key elements that explain the fundamental difference between classical physics and quantum physics. Moreover, the possibility to take advantage of such concepts to process information is the basis of a new approach to information, quantum information science, where they also imply a difference in nature between classical and quantum information. The theoretical and experimental ability to process information encoded on particles that obey quantum physics gave birth in the 1980's to new areas of research such as quantum computing, first imagined by Richard Feynman, or quantum cryptography, first proposed by Charles Bennett and Gilles Brassard.

The objective of quantum information theory is to understand what can and cannot be done, from the point of view of information sciences, with quantum particles. In this thesis, we focus on one informational concept, randomness, and one quantum feature, Bell nonlocality, and we study their relations.

1.2 Motivation and main contributions

Understanding what is random is both a theoretical and a practical question. From a foundational point of view, the inherently random nature of the laws of quantum physics is perplexing, as it contrasts with a long-standing view of what a physical theory should be: the goal of classical physics was to provide deterministic explanations for observed phenomena. From a practical point of view, the ability to generate random numbers is crucial, as it is a pre-requisite for several information tasks. Defining adequate measures of randomness is thus crucial for quantum information theory.

The notion of randomness can be related to the concept of nonlocality, which concerns the correlations between events observed by distant agents. The nonlocal character of said correlations does not depend on the underlying physical system: it can thus be witnessed even when the physical systems from which it emerges are not known or not perfectly characterised. Studying randomness

though the prism of nonlocality thus allows to quantify randomness without relying on a physical description of the underlying systems. This statement has two major consequences. Conceptually, it implies that the random nature of a physical process can be assessed without relying on a physical theory, and, in particular, on quantum physics. Practically it means that one can certify that a process generates random numbers in a device-independent way, that is, without knowing the physical systems that underlie that process. We now describe the various scenarios in which we derived trade-offs between randomness and nonlocality.

1.2.1 Randomness based on time-ordering and no-signalling

One can certify, in a device-independent way, that a physical process produces random numbers, even without relying on the quantum formalism, as long as one assumes that the 'no-signalling' principle holds, that is, that information cannot be instantaneously transmitted. If that process is nonlocal, its outputs cannot be perfectly predicted by a third party. One can thus quantify its randomness by estimating how well the outputs can be predicted by said third party. However, that measure is typically defined for only one instance of the process. If the process is repeated several times and if one wants to derive, from this one-round measure, a trade-off between non-locality and randomness that would hold for all the repetitions, additional assumptions have to be made: one could assume, e.g., that the repetitions are independent and identically distributed (i.i.d.), or that they are causally independent.

Contribution We define a measure of unpredictability that directly takes into account the repetitions of the process. We present different causal structures to model how these repetitions are related to each other. We show that, if one assumes that past events can influence future events, but not the converse ('time-ordering'), the unpredictability per repetition decreases at each repetition. This result might help in understanding whether privacy amplification based on no-signalling only is possible, that was proven to be impossible if no time structure is assumed [HRW13] but remains open otherwise [AFTS12].

1.2.2 Practical randomness

When one derives randomness versus nonlocality trade-offs, one can quantify how nonlocal a process is via the evaluation of a Bell expression. The amount of

randomness that is certified depends on the Bell expression that is evaluated. For a given process, there exists an optimal Bell expression, i.e., one that certifies the maximal amount of randomness generated by the process. That expression can be easily derived when the correlations governing the process are known and obey the no-signalling principle. However, in a practical situation where one wants to proceed in a device-independent way, those correlations are not accessible. One can only estimate them by repeating the process several times and collecting the frequencies. However, due to finite statistics, the correlations derived from the frequencies do not obey no-signalling, even if the actual underlying correlations do. In that case, no optimal Bell expression can be derived from it, as the notion of nonlocality-based randomness is only relevant in a no-signalling world.

Contribution We propose to use a regularisation method [LRZ+18] that projects the frequencies collected from a practical situation onto the space of correlations that obey the no-signalling principle. That enables us to derive a Bell expression that is well suited for a specific process. We can then derive a lower-bound on the unpredictability of the data outputted by the process as a function of the estimated value of that Bell expression, using known techniques [NSBSP18]. Numerical simulations show the efficiency of our method.

1.2.3 Randomness in a tri-partite scenario

Trade-off between randomness and nonlocality are generally derived in a scenario involving two agents, the minimal requirement for the notion of nonlocality to make sense. However, Bell nonlocality, as well as device-independent randomness, can be defined for more than two parties. The questions of deriving trade-offs for three parties or more and finding applications of such trade-offs to new device-independent information protocols haven't been explored.

Contribution We quantify how much randomness can be certified when three parties evaluate the violation of the tri-partite Mermin inequality [Mer90]. We give the analytical values of the unpredictability contained in the parties' outcomes, as a function of the Mermin inequality violation, by deriving sum-of-square (SOS) decompositions [BP15]. We then discuss the possibility of using these results to design a multi-partite protocol, namely, secret sharing, in a device-independent manner, and argue that it seems unlikely to work.

1.2.4 Maximal randomness from partially entangled states

The amount of randomness that can be certified from a given process has a theoretical maximal value, which corresponds to the case where the uncertainty on the outcomes is maximal for a third party. For instance, a process that outputs two bits can produce at most two bits of randomness, which is guaranteed when a third party cannot predict these outputs better than with a uniform guess. Such a process is actually realised with a physical system, namely, a quantum state on which measurements are performed. In the device-independent approach, they are treated as black boxes: the trade-off between nonlocality and randomness is evaluated for a given process, and is then valid independently of the physical system. However, there might exists requirements on the underlying state and measurements for a given process to be achieved. For instance, can any entangled state give rise to a process that certifies maximal randomness?

Contribution We show that, for a process with two dichotomic outputs, the maximal value of two bits of randomness can be certified from any entangled two-qubit pure state. The fact that randomness, nonlocality and entanglement are inequivalent quantities was already observed in [AMP12], where it was shown that almost maximal randomness could be certified from almost unentangled pure states. Though entanglement is necessary for certifying device-independent randomness, our result now shows that, for pure states, the amount of randomness and the level of entanglement are completely uncorrelated quantities.

Chapter 2

Preliminaries

In this section, we introduce the various concepts that we will use in this thesis. We first describe the notion of Bell nonlocality. We then define the device-independent approach to quantum information. Lastly, we present some notions related to the concept of randomness.

2.1 Bell nonlocality

We start by describing the setting of a Bell test and we introduce the corresponding notations. We then give a characterisation of the local, quantum and no-signalling sets. We conclude by defining the notion of Bell inequality. Though we take a historical approach, all concepts are presented in modern phrasing.

2.1.1 Alice and Bob

In order to present the key concepts and the main results of this thesis, we will refer on numerous occasions to two agents, Alice and Bob. This convention comes from the field of classical cryptography: they were first mentioned in a 1978 article by Ron Rivest, Adi Shamir and Leonard Adleman [RSA78], to refer to two distant agents who aim to communicate securely, and who were previously referred to as 'A' and 'B'.

In the context of quantum information theory, Alice and Bob are two fictional, possibly distant observers, each interacting with a physical, possibly quantum system (see Fig. 2.1). Alice interacts with her system $\mathcal{A}$ by choosing an input

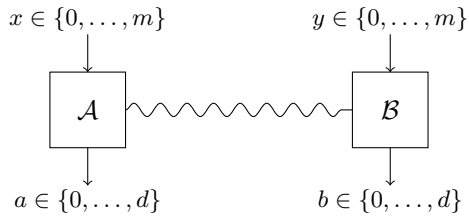


Figure 2.1: Black box representation of a Bell test: Alice and Bob respectively input x and y into two uncharacterised devices $\mathcal{A}$ and $\mathcal{B}$, and respectively obtain outputs a and b. The curvy line connecting their devices represent possible correlations between $\mathcal{A}$ and $\mathcal{B}$ due, for instance, to entanglement.

labeled $x \in \{0, \ldots, m\}$ and obtaining an output labeled $a \in \{0, \ldots, d\}$. Bob does the same respectively with $\mathcal{B}$, $y \in \{0, \ldots, m\}$ and $b \in \{0, \ldots, d\}$. The inputs can be thought of as measurement choices and the outputs as measurement results. We call this experiment 'a Bell test'. If the Bell test is repeated several times, we refer to a single interaction as 'a run' or 'a round'. In this thesis, we restrict ourselves to the case of dichotomic measurements, i.e., $d = 1$. We sometimes label '-1' and '$+1$' the possible outcomes a and b for convenience. We always specify it if we do so.

We always denote the random variable associated to a value by capitalising it: A, B, X, Y correspond to the random variables associated to the possible choices a, b, x, y, respectively, and we denote $P_{AB|XY}(ab|xy)$ the conditional probability of obtaining the output pair (a, b) when the pair (x, y) was inputed. Unless there is some ambiguity, we will omit the corresponding random variables in the rest of this thesis and write $P(ab|xy)$ only. From now on, we refer to $P(ab|xy)$ as 'the underlying distributions', 'the correlations' or 'the behaviour'. This description is often referred to as a 'black box' approach, because no mention of the actual physical set-up from which these correlations arise is needed.

The correlations $P(ab|xy)$ can be studied as a pure mathematical object, as is done in part of this thesis. The question of its relation to practical situations is however not trivial. The simplest approach is to consider that the random variables A, B, X, Y behave in an independent and identically distributed (i.i.d.) way every time that Alice and Bob interact with the systems $\mathcal{A}$ and $\mathcal{B}$. In that case, by repeating the Bell test several times, one can compute the frequencies of all input-output pairs, and use them as an estimate of the underlying distribution $P(ab|xy)$.

In some parts of this thesis, we will not make the assumption that the random variables associated to each Bell test are i.i.d.. In that case, if the Bell test is repeated n times, for some round $i \in [n]$, we will write (a_i, b_i, x_i, y_i) the inputs and outputs associated to this round i, (A_i, B_i, X_i, Y_i) the corresponding random variables, and $P_{A_iB_i|X_iY_i}(a_ib_i|x_iy_i)$ (or $P(a_ib_i|x_iy_i)$) the corresponding underlying conditional distributions. We will then denote sequences of inputs, outputs, and random variables in bold font: $\mathbf{x} = (x_1, ..., x_n)$ and similarly for $\mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{X}, \mathbf{Y}, \mathbf{A}$ and $\mathbf{B}$.

When we introduce a quantity defined for a given random variable and evaluated on its associated probability distribution, we sometimes don't make the dependence of that quantity on the distribution explicit, if we consider only one distribution for that given random variable.

2.1.2 The local, quantum and no-signalling sets

In a 1935 paper [EPR35], Albert Einstein, Boris Podolsky and Nathan Rosen designed a thought experiment in which two distant agents (precursors of Alice and Bob) perform some measurements on a quantum system that consists of two entangled particles. They felt uneasy with the correlations that arose from this experiment (which would be called a Bell test a few decades afterwards) and claimed that what would later on become the EPR paradox showed that quantum theory was incomplete.

Before we address this point, let us look into the possible sets of behaviours $P(ab|xy)$ that Alice and Bob can obtain when they perform a Bell test.

If we accept the validity of the quantum formalism, we would then consider that the correlations between Alice and Bob's inputs and outputs arise from performing measurements on a quantum state. Formally, Alice and Bob share a state ρ_{AB} that belongs to a joint Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, on which Alice (resp. Bob) performs the measurements $\{M^A_{a|x}\}$ (resp. $\{M^B_{b|y}\}$).

Note that we don't lose any generality by assuming that the state is pure and that the measurements are projective: since there is no restriction on the dimension of the Hilbert space, we can always see mixed states and positive-operator valued measures (POVM) as pure states and projective measurements in a Hilbert space of higher dimension.

Applying the Born rule, we can now characterise the set of quantum behaviours.

Definition 1. *We say that a behaviour $P(ab|xy)$ is quantum, and we write $P \in \mathcal{Q}$, if there exists a pure state $|\Psi_{AB}\rangle$ and projective measurements $\{M^A_{a|x}, M^B_{b|y}\}$*

such that

$$P(ab|xy) = \langle \Psi_{AB}| \, M^A_{a|x} \otimes M^B_{b|y} \, |\Psi_{AB}\rangle \, . \tag{2.1}$$

Another model for $P(ab|xy)$ arises if we consider that the behaviour observed by Alice and Bob should be explainable separately for Alice on one side and for Bob on the other side. In that case, the behaviour should be decomposable into a probability response function for Alice, that describes the relation between her output and her input, and a similar probability response function for Bob. This approach doesn't prevent us from taking into account some shared randomness between Alice and Bob, i.e., some (possibly hidden) variables $\lambda \in \Lambda$ that are distributed by some common source to Alice and Bob according to a distribution $q(\lambda)$, and have an influence on Alice and Bob's response functions. That characterises the local set.

Definition 2. *We say that a behaviour $P(ab|xy)$ is local, and we write $P \in \mathcal{L}$, if there exists some variables λ distributed according to $q(\lambda)$ and some probability distributions $\{P(a|x,\lambda),\ P(b|y,\lambda)\}$ such that*

$$P(ab|xy) = \int_\Lambda q(\lambda) P(a|x,\lambda) P(b|y,\lambda) d\lambda. \tag{2.2}$$

Lastly, one can consider that the only constraints on the correlations between Alice and Bob should be that they don't allow for instantaneous signalling. Indeed, in a Bell test, the two parties can be arbitrarily distant. If Bob's marginal distributions depend on Alice's choice of inputs, Alice could use her choice of input to instantaneously transmit information to Bob. This faster-than-light transmission would be in conflict with relativity. The set of behaviours preventing this is defined in the following way.

Definition 3. *We say that a behaviour $P(ab|xy)$ is no-signalling, and we write $P \in \mathcal{NS}$, if the following relations are satisfied:*

$$\forall\ a, x, y, y' \ \sum_b P(ab|xy) = \sum_b P(ab|xy') \triangleq P(a|x), \tag{2.3}$$

$$\forall\ b, y, x, x' \ \sum_a P(ab|xy) = \sum_a P(ab|x'y) \triangleq P(b|y). \tag{2.4}$$

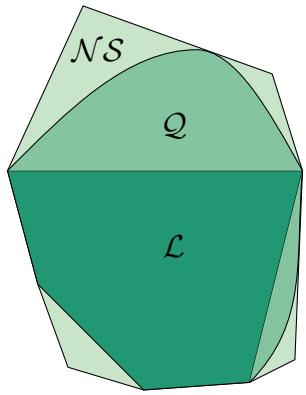


Figure 2.2: Sketch of the local, quantum and no-signalling sets that represents the strict inclusions $\mathcal{L} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}$, as well as the fact that $\mathcal{L}$ and $\mathcal{NS}$ are polytopes.

A modern (and somewhat simplified) reading of the EPR paradox could be as follow: Einstein, Podolsky and Rosen wanted $\mathcal{L}$ and $\mathcal{Q}$ to coincide, because they felt that only $\mathcal{L}$ was an admissible description of a physical theory. They thus believed that quantum physics could and should be completed in such a way that all quantum correlations admit a local hidden variable model. Moreover, Einstein did not accept that the laws of physics could be probabilistic: the hidden variables should represent our ignorance about quantum theory. Once added to the theory, they would render it not only local but also deterministic — these two notions being equivalent, as we explain in the next section.

This intuition proved to be wrong a few years later. Before we go into details about that, we finish this section by giving a few properties of these three sets (see [BCP$^+$14] for a detailed review). They are closed, bounded and convex, and obey the following inclusion relation:

$$\mathcal{L} \subsetneq \mathcal{Q} \subsetneq \mathcal{NS}. \tag{2.5}$$

The local and no-signalling sets are polytopes, and can thus be characterised as the convex hull of a finite number of extremal points, the vertices, or as the interior of a finite number of hyperplanes, the facets. The quantum set is not a polytope and is much harder to characterise. A sketch of these three sets that illustrates these properties can be found in Figure 2.2.

Some results of this thesis rely on the validity of quantum physics (i.e., holds

for $P(ab|xy) \in \mathcal{Q}$), others only on the no-signalling principle (i.e., holds for all behaviours $P(ab|xy) \in \mathcal{NS}$ that can even be 'supra-quantum').

2.1.3 Bell inequalities

In 1964, John Stewart Bell published an article on the EPR paradox [Bel64]. While general physical principles like 'local realism' were formerly not properly defined, Bell gave a precise mathematical characterisation of such concepts. This enabled him to obtain the following result:

Theorem 1. *No physical theory of local hidden variables can ever reproduce all the predictions of quantum mechanics.*

To prove this result, Bell derived an inequality that has to be satisfied by any local correlations (as defined in Definition 2), but that some quantum correlations do not satisfy. Namely, he exhibited a hyperplane that separates a quantum behaviour from the whole set of local behaviours. The strength of this result resides in the fact that it provides a simple mathematical quantity that, on the one hand, forbids irremediably any local explanation for the set of quantum correlations, and, on the other hand, can be experimentally tested.

Many such inequalities were then derived, and were termed 'Bell inequalities'. Deriving such inequalities can now be achieved automatically, as they correspond to the facets of the local polytope. However, for a given number of inputs and outputs, the corresponding local polytope is usually characterised by its vertices. Determining the facets of a polytope, given its vertices, is called a convex hull problem, and can be prohibitively time consuming.

The generic form of a Bell inequality is an inequality that is linear in $\{P(ab|xy)\}$:

$$\sum_{a,b,x,y,} c_{abxy} P(ab|xy) \triangleq \mathcal{I}(P(ab|xy)) \leq I \tag{2.6}$$

We'll refer to $\mathcal{I}(P(ab|xy))$ as a 'Bell expression' and to $\mathcal{I}(\cdot)$ as the corresponding Bell functional. The maximal I that can be obtained for $P(ab|xy) \in \mathcal{L}$ is called 'the local bound', and for $P(ab|xy) \in \mathcal{Q}$, 'the quantum bound' or 'Tsirelson bound'.

The simplest Bell inequality was exhibited by John Clauser, Michael Horne, Abner Shimony and Richard Holt in 1969 [CHSH69] and is abbreviated as 'CHSH inequality'. It is defined in the scenario where Alice and Bob both have two measurements choices, each of them having two possible measurements results. It reads:

$$S \triangleq \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2, \tag{2.7}$$

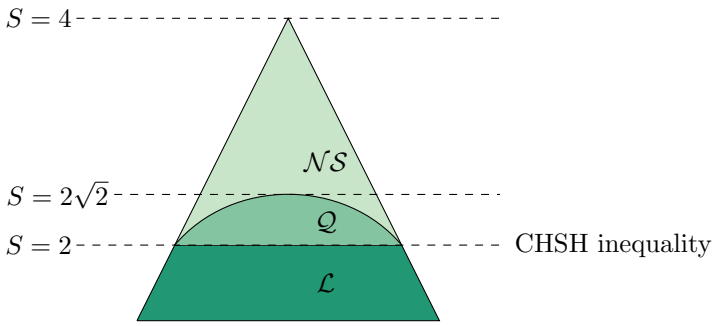


Figure 2.3: Sketch of the CHSH inequality as a hyperplane dissociating the local set from some quantum and no-signalling behaviours.

where the correlator $\langle A_x B_y \rangle$ is equal to:

$$\langle A_x B_y \rangle \triangleq \sum_{a,b} (-1)^{a+b} P(ab|xy). \tag{2.8}$$

If Alice and Bob share the Bell state:

$$|\Phi^+\rangle \triangleq \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{2.9}$$

on which they perform the measurements described by the following observables:

$$\begin{aligned} A_0 &= \sigma_{\mathrm{z}}, & B_0 &= \frac{\sigma_{\mathrm{z}} + \sigma_{\mathrm{x}}}{\sqrt{2}}, \\ A_1 &= \sigma_{\mathrm{x}}, & B_1 &= \frac{\sigma_{\mathrm{z}} - \sigma_{\mathrm{x}}}{\sqrt{2}}, \end{aligned} \tag{2.10}$$

where σ_{x} and σ_{z} are the X and Z Pauli matrices, they will then obtain $S = 2\sqrt{2}$. This value is actually the highest achievable by a quantum behaviour [Tsi80]. Note that a supra-quantum behaviour, the Popescu-Rohrlich (PR) box [PR94], defined as:

$$P(ab|xy) = \frac{1}{2} \delta_{a \oplus b, x \cdot y} \tag{2.11}$$

can achieve the maximal value $S = 4$ (see Fig. 2.3).

As previously mentioned, the quantity S can be experimentally evaluated. This was achieved in a convincing manner for the first time by the team of Alain

Aspect, in a series of experiments performed between 1980 and 1982 [AGR82], where they indeed observed $S > 2$. Experimentally meeting all the requirements of a Bell test is challenging, as several loopholes can be caused by the set-up. A lot of work was thus still devoted to obtaining more and more experimental evidences that Bell inequalities could be violated. The first loophole-free Bell inequality violation was reported in 2015 [HBD+15].

This sequence of experimental validations of Bell's groundbreaking result provided a definitive answer: some correlations observed in nature cannot be explained by a local hidden variable model. We thus term them 'nonlocal'. Far from being an isolated and purely theoretical discovery, this proved to be an extremely fruitful source of results in the field of quantum information science.

2.2 Device-independent quantum information theory

We first describe the essence of the link between nonlocality and information. We then explain how that link can be used for designing information protocols in a device-independent way, and we explain how uncharacterised quantum resources can be characterised only via their associated behaviours.

2.2.1 Characterising nonlocal correlations

After the set of local behaviours was mathematically characterised, properties inherent to local and nonlocal correlations could be demonstrated. In 1982, Arthur Fine derived the following result [Fin82] (we use the exact phrasing of the original article, 'the experiment' should be understood, in our terminology, as 'the Bell test'):

Theorem 2. *The following statements are equivalent:*

(i) There is a deterministic hidden-variables model for the expriment.

(ii) There is a factorisable, stochastic model.

(iii) There is one joint distribution for all observables of the experiment, returning the experimental probabilities.

(iv) There are well-defined, compatible joint distributions for all pairs and triples of commuting and non commuting observables.

(v) The Bell inequalities hold.

Point (ii) corresponds to our definition of the local set $\mathcal{L}$, point (v) thus follows from the way we defined Bell inequalities. The most important point for this thesis is (i): it states, in essence, that local is equivalent to deterministic. More precisely, it states that a behaviour is local if and only if there exists some variables that, when added to the model, makes it deterministic. These variables might be unknown for now, hence the term 'hidden', yet they exist: the apparent randomness in the results of the Bell test is merely due to our ignorance.

That implies in return that no such deterministic explanations exist for a nonlocal behaviour: the probabilistic nature of the input-output correlations is intrinsic. In other words, any correlations that violates a Bell inequality is inherently random.

From a fundamental point of view, it represents a milestone in our understanding of physics: while physicists used to be concerned with deriving deterministic explanations of the phenomena observed in nature, it was proven that the statistics observed in a Bell test could never be explained in a deterministic manner.

From an applied point of view, it constitutes a fertile ground for information processing and communication: the ability to generate random bits, and to certify their random character, with no need to rely on a specific description of natural processes, paved the way to a new approach to information theory.

Before we move on to presenting the possible applications of nonlocality to information protocols, let us mention an obstacle that we have to face when we wish to evaluate a information-theoretic figure of merit (see Section 2.3 for details), not on the whole space of nonlocal behaviours $\mathcal{NS}$, but specifically on the space of quantum behaviours $\mathcal{Q}$. While $\mathcal{NS}$ is easy to characterise, $\mathcal{Q}$ has a complex mathematical structure. A simple characterisation of it would yet be very valuable, both from fundamental and applied perspectives. On the one hand, it would enable us to understand what singles out quantum physics amongst other nonlocal theories. On the other hand, optimising various quantities over the set of all quantum behaviours is necessary for numerous information tasks allowed by the nonlocal nature of quantum physics, as we will see later on. However, this can only be done with a computationally efficient characterisation of $\mathcal{Q}$.

Such a characterisation remains to be found. Nevertheless, in 2007, an outer approximation of the quantum set based on semi-definite programming was presented in [NPA07, NPA08]. It consists of a hierarchy of sets $\{\mathcal{Q}_k\}_{k=1}^{\infty}$, later on called 'the NPA hierarchy', that converges to the quantum set $\mathcal{Q}$ (cf

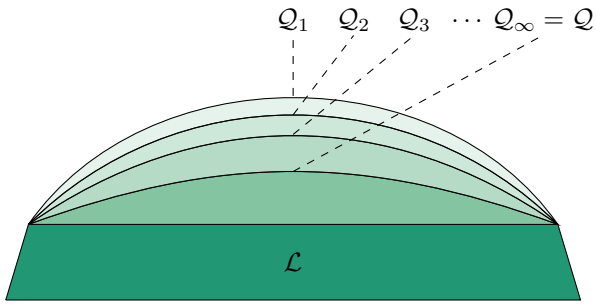


Figure 2.4: Outer approximation of the quantum set by the NPA hierarchy.

Fig. 2.4). It is based on constructing moment matrices that are constrained by semi-definite positiveness. It provides a practical approximation of the quantum set, see Section 2.3.1 for examples of application.

We introduced the idea that nonlocal behaviours have some inherent properties, i.e., properties that hold independently of how these correlations were obtained. This is the core idea of the Device-Independent (DI) approach to quantum information, that we now elaborate.

2.2.2 Nonlocal correlations for cryptographic protocols

Most of the cryptographic protocols in use today are based on mathematical conjectures. For instance, the communication between two distant agents, Alice and Bob, is undecipherable provided that any potential eavesdropper, hereafter called 'Eve', is unable to solve some mathematical problem that we believe too hard to be solved in a reasonable time. In 1984, Charles Bennett and Gilles Brassard proposed a key distribution protocol where some information sent by Alice to Bob was encoded in the state of a quantum particle, say, the polarisation of a photon [BB84]. By basing the security of what would then become the BB84 protocol on the laws of quantum mechanics, and in particular on the no-cloning theorem, they could design a scheme for key distribution that was 'unconditionally' secure. This was the birth of Quantum Key Distribution (QKD).

However, one should bear in mind that the word 'unconditionally' here refers to the fact that the security is not based on the unproven hardness of a computational problem. But ultimately, the security of any cryptographic protocol relies on some assumptions, and is, in that sense, conditional. Indeed,

in 2006, a team of researchers reported that they had hacked a cryptographic system that implemented the BB84 protocol [MAS06]. To do so, they used the mismatch between the theoretical requirements of the BB84 protocol and its experimental realisation: only if the protocol had been perfectly implemented would the theoretical security proof apply and would such hacking be impossible. Even though quantum physics exhibits nonclassical features that enable us to process information in a revolutionary way, implementing quantum information protocols is hard because controlling accurately enough quantum systems is extremely difficult.

One of the aims of the device-independent approach to quantum information protocols is precisely to circumvent this challenge. The core idea, mentioned in the previous section, is the following: a nonlocal behaviour cannot have a deterministic explanation, and this intrinsic randomness can be put to use for several information-theoretic tasks. Even if we are in complete ignorance of the internal working of the devices, and even if we are facing an omniscient and omnipotent eavesdropper Eve, the nonlocal character of a behaviour guarantees that it produces outputs that are (at least partially) unpredictable to Eve, whatever possible backdoors she could use and whatever possible hidden variables she could know. Moreover, the nonlocality of a behaviour can be easily experimentally witnessed by, e.g., the observation of a Bell inequality violation.

We now give a brief overview of which cryptographic protocols were explored in the device-independent framework, see [PSV16] for a recent state-of-the-art. By cryptographic protocol, we mean a process that aims to create or keep some information secret and safe. Examples of such tasks are:

- Randomness Generation (RNG)
- Key distribution
- Bit commitment
- Secret sharing
- Authentification

In this thesis, we focus on the first one, but our results have implications for others, in particular for key distribution. The goal of RNG is to obtain bits that are secret, i.e., that cannot be predicted by an adversary, in one location. This is cryptographic primitive that is useful for many tasks, such as the ones listed above. The goal of QKD is to obtain two identical sequences of secret bits in two

distant locations. This then allows two distant agents to communicate securely by using one-time pad.

One usually considers that the first task that was achieved in a device-independent manner was QKD, with a protocol based on entanglement defined by Artur Ekert in 1991 [Eke91]. At that time, however, the device-independent potential of this protocol was not truly apprehended. In 1998, Dominic Mayers and Andrew Yao proposed a QKD protocol based on the concept of 'self-checking source' [MY98]: without calling it that way, they understood the possibility of performing QKD in a device-independent manner using nonlocality. In 2005, Jonathan Barrett, Lucien Hardy and Adrian Kent realised that key distribution could be achieved based on the sole assumption of no-signalling, i.e., without relying on the validity of quantum physics [BHK05]. In 2006 Roger Colbeck showed that this was also true for RNG [Col06]. Finally, the potential of using nonlocality to design device-independent protocols was understood by Antonio Acín *et al.* for QKD in 2007 [ABG$^+$07], and by Stefano Pironio *et al.* for RNG in 2010 [PAM$^+$10], where the term 'device-independent', along with the first security proofs based on the observation of a Bell inequality violation, were introduced.

In what we just described, the observation of nonlocality enables us to derive information-theoretic figures of merit related to RNG and QKD. These derivations can be based of the validity of quantum physics, but don't have to be: they relate nonlocality and cryptographic tasks. We now describe how nonlocality can be used for certifying the quantum character of a Bell test.

2.2.3 Nonlocal correlations for resource certification

In 2004, Mayers and Yao defined the concept of 'self-testing' [MY04]: in the framework of quantum physics, some behaviours $P(ab|xy)$ can only be achieved by performing essentially unique quantum measurements on a unique quantum state. When one is given two black boxes that perform a Bell test, if one observes such a behaviour, one can then be sure that these black boxes consist of that specific state and measurements: they are not black boxes anymore.

Let us make this statement precise. We call $\{A'_x\}$ and $\{B'_y\}$ the (uncharacterised) observables that the black boxes perform, and $|\Psi'\rangle \in \mathcal{H'}_A \otimes \mathcal{H'}_B$ the (uncharacterised) state on which they operate: $\{|\Psi'\rangle, A'_x, B'_y\}$ is called the physical experiment. We then define a reference experiment $\{|\Psi\rangle, A_x, B_y\}$, that consists of well characterised state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and observables $\{A_x, B_y\}$.

Definition 4. *We say that the physical experiment is equivalent to the reference*

experiment if there exists a local isometry $\Phi = \Phi_A \otimes \Phi_B$*, with* $\Phi_A : \mathcal{H'}_A \to \mathcal{H'}_A \otimes \mathcal{H}_A$ *and* $\Phi_B : \mathcal{H'}_B \to \mathcal{H'}_B \otimes \mathcal{H}_B$*, such that:*

$$\Phi(|\Psi'\rangle) = |junk\rangle \otimes |\Psi\rangle\,, \tag{2.12}$$

$$\Phi(A'_x \otimes B'_y\,|\Psi'\rangle) = |junk\rangle \otimes (A_x \otimes B_y\,|\Psi\rangle). \tag{2.13}$$

It means that the reference experiment is an accurate description of the devices, up to local unitaries and irrelevant degrees of freedom that are represented by the additional 'junk' state. Let us however add that this definition doesn't encompass some transformations that are not visible from the behaviour only and not physical, such as complex conjugation. However, certifying a state and some measurements only up to, say, complex conjugation, is often sufficient for applications such as information protocols.

For a physical experiment and a reference experiment to be equivalent, it is necessary that their associated correlations are the same. Finding a self-test for a given behaviour $\{P(ab|xy)\}$ amounts to proving that it is also sufficient. In some cases, the complete description of a behaviour is not even needed: the value of a Bell expression can be sufficient.

For example, we mentioned in Section 2.1.3 that the state given in Eq. (2.9) and the measurements described by Eq. (2.10) yield $S = 2\sqrt{2}$. The reverse is also true: the observation of correlations between two black boxes such that $S = 2\sqrt{2}$ certifies that the physical experiment is equivalent to the reference experiment $\{|\Psi\rangle, A_0, A_1, B_0, B_1\}$ with:

$$|\Psi\rangle = |\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \qquad \begin{aligned} A_0 &= \sigma_{\mathrm{z}}, & B_0 &= \frac{\sigma_{\mathrm{z}} + \sigma_{\mathrm{x}}}{\sqrt{2}}, \\ A_1 &= \sigma_{\mathrm{x}}, & B_1 &= \frac{\sigma_{\mathrm{z}} - \sigma_{\mathrm{x}}}{\sqrt{2}}. \end{aligned} \tag{2.14}$$

In order to construct a self-test for given state and measurements, the corresponding behaviour must be extremal in $\mathcal{Q}$: if it is not extremal, it can be decomposed as various mixtures of other behaviours, and thus can be obtained with various sets of state and measurements.

A Sum-Of-Squares (SOS) decomposition is a useful tool to prove an equivalence relation based on the observation of the maximal value of a Bell expression. For a given Bell operator $\mathcal{I}$ with maximal quantum value I_Q, the following holds:

$$\forall\,|\Psi\rangle\,, \langle\mathcal{I}\rangle_\Psi = \langle\Psi|\,\mathcal{I}\,|\Psi\rangle \leq I_Q \Leftrightarrow I_Q \cdot \mathbb{1} - \mathcal{I} \succeq 0. \tag{2.15}$$

Since the operator $I_Q \cdot \mathbb{1} - \mathcal{I}$ is semi-definite positive, one can try to decompose it as a finite sum of squares:

$$I_Q \cdot \mathbb{1} - \mathcal{I} = \sum_{i=1}^{s} P_i^\dagger P_i, \tag{2.16}$$

where each P_i is a polynomial function in $\{A_x\}$ and $\{B_y\}$. When evaluated on the state $|\Psi\rangle$ that maximally violates the corresponding Bell inequality, this expression yields, in turn:

$$(I_Q \cdot \mathbb{1} - \mathcal{I})\,|\Psi\rangle = 0 \Leftrightarrow \forall i \in \{1, .., s\}, P_i\,|\Psi\rangle = 0. \tag{2.17}$$

This set of equations can then be used to derive the self-testing equivalence relation. This was for instance done in [BP15], where the authors derived a self-test for all partially entangled two qubit pure states, i.e., using the Schmidt decomposition, all states of the form:

$$|\Psi_\theta\rangle \triangleq \cos(\theta)\,|00\rangle + \sin(\theta)\,|11\rangle \tag{2.18}$$

for $\theta \in [0, \pi/4]$.

Note that the nonlocal character of a behaviour is needed neither to define the notion of self-test nor to prove an equivalence relation. For the latter, only the extremality in the quantum set is a pre-requisite. However, from a quantum information-theoretic point of view, self-testing local correlations is fruitless: such correlations cannot provide any quantum advantage for information processing. On the contrary, self-testing a nonlocal quantum behaviour is useful: it certifies that some a priori black boxes actually contain specific measurements and state in a device-independent way. These black boxes can then be used to run a cryptographic protocol that is described in terms of said measurements and state, as was proposed in [MY04].

Being able to certify the quantum description of a behaviour from purely classical observation has a fundamental interest. It is also useful from a practical point of view, as it enables us to run device-independent cryptographic protocols. However, RNG need not be based on self-testing to be achieved in a device-independent way. We now explain how to certify and quantify the randomness produced by a Bell test.

2.3 Randomness

We first introduce the concept of guessing probability associated to a Bell test. We then explain how it can be seen as a quantity of fundamental interest, and

how it can be used as a building block for device-independent cryptographic protocols.

2.3.1 The guessing probability

The guessing probability quantifies the ability to predict the outcome of a probabilistic process. This broad notion admits several mathematical definitions, depending on the information that is accessible and on the physical theory that underlies the process.

Our framework is that of device-independent cryptography: the situation is thus modelled by an adversarial black box scenario, where Eve tries to guess some outcomes obtained by Alice and Bob via a process described by the behaviour $P(ab|xy)$. If Eve knows the underlying distributions, but cannot tamper with them, the optimal strategy for predicting the output pair associated to a given input pair (x, y) consists in guessing the most probable one. Writing G_{xy} the corresponding guessing probability, we get:

$$G_{xy} = \max_{ab} P(ab|xy). \tag{2.19}$$

However, in an adversarial scenario, we want to take a conservative approach: what if Eve had prepared the black boxes that Alice and Bob use? Is there then any way that Alice and Bob obtain outputs that are unknown to Eve? If no contraints are imposed on the underlying distributions, that is of course impossible: Eve could prepare deterministic devices. However, if the black boxes are constrained to violate a Bell inequality, it is possible, due to Theorem 2. In that case, we define the guessing probability as [PAM+10]:

$$\begin{aligned} G_{xy} = \max_{ab} \max_{P} \quad & P(ab|xy), \\ \text{s.t.} \quad & \mathcal{I}(P) = I, \\ & P \in \mathcal{Q}. \end{aligned} \tag{2.20}$$

It operationally corresponds to the case where Alice and Bob have no control over the devices, which enables Eve to design them in the most favourable way possible, but where Alice and Bob check that a given Bell inequality, described by $\mathcal{I}$, is violated with value I. That is enough to certify the presence of some randomness, and constitutes the basic idea of the device-independent paradigm.

The second constraint, which imposes that the behaviour has a quantum origin, can be relaxed: one can impose that the behaviour belongs to one of the

NPA hierarchy sets $\mathcal{Q}_k$, or to the no-signalling set $\mathcal{NS}$. From a practical point of view, these changes make G_{xy} efficiently computable, and the solutions of these modified versions of (2.20) are upper bounds on the quantum guessing probability. From a fundamental point of view, imposing that the behaviour lies in $\mathcal{NS}$ instead of $\mathcal{Q}$ corresponds to assuming that Eve is 'supra-quantum': she can prepare distributions that are not accessible via quantum physics.

Before we go further into details about the guessing probability, let us introduce some terminology related to optimisation problems such as (2.20). The function that is optimised is called the objective function. The constraints define a subspace of $\mathbb{R}^{(d+1)^2(m+1)^2}$, called the feasible region. A point (in our case, a behaviour) that belongs to the feasible region is called a feasible point. When one imposes $P \in \mathcal{Q}_k$, the feasible region is the intersection of the cone of positive semidefinite matrices with an affine space, i.e., a spectrahedron; while the feasible region associated to $P \in \mathcal{NS}$ is the intersection of a finite number of half spaces, i.e., a convex polytope. The optimisation of a linear function over a spectrahedron is called Semidefinite Programming (SDP), and Linear Programming (LP) when it is over a convex polytope. Both LP and SDP can be efficiently numerically solved.

The problem introduced in (2.20) is a measure of the unpredictability of the output pair (a, b). One can also define the guessing probability associated to one output, say, a, in the following way:

$$\begin{aligned} G_x = \max_a \max_P \quad & P(a|x), \\ \text{s.t.} \quad & \mathcal{I}(P) = I, \\ & P \in \mathcal{Q}. \end{aligned} \tag{2.21}$$

Eq. (2.20) defines the 'global' guessing probability, whereas Eq. (2.21) corresponds to the 'local' guessing probability. In the case of binary outputs, the local guessing probability can vary from 1 (complete predictability) to 1/2 (complete unpredictability), and the global guessing probability from 1 to 1/4.

In [PAM+10, MPA11], the authors proved that, when $\mathcal{I}$ is the CHSH expression defined in (6.4), the local quantum guessing depends on the CHSH value $I = S$ as:

$$G_x(S) = \frac{1}{2}(1 + \sqrt{2 - S^2/4}) \tag{2.22}$$

and the local no-signalling guessing probability as:

$$G_x(S) = \frac{3}{2} - \frac{S}{4} \tag{2.23}$$

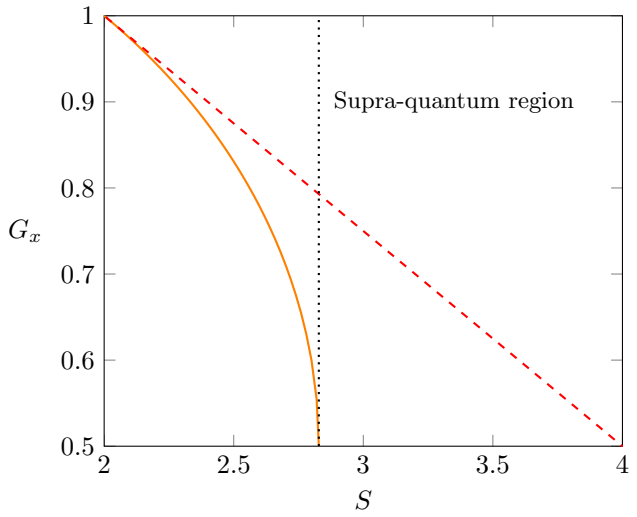


Figure 2.5: Local guessing probability as a function of the CHSH violation β. The orange solid line corresponds to the quantum guessing probability, the dashed red line to the no-signalling one. For the local bound ($\beta = 2$), the predictability is maximal, while for the maximal CHSH values, the unpredictability is maximal.

These values are plotted in Fig. 2.5.

The definition of the guessing probability given in Eqs. (2.20) and (2.21) is however not satisfactory if the obtained result is not concave in I. Indeed, let's suppose that, for some values I_1 and I_2 associated to a Bell expression $\mathcal{I}$, and for some $\mu \in [0, 1]$, the following holds:

$$G_{xy}\big(\mu \cdot I_1 + (1-\mu) \cdot I_2\big) \leq \mu \cdot G_{xy}\big(I_1\big) + (1-\mu) \cdot G_{xy}\big(I_2\big). \tag{2.24}$$

Then, by mixing her preparations for I_1 (with weight μ) and I_2 (with weight $1-\mu$), Eve would improve her guess, while still satisfying the constraint on the Bell expression value, as it is linear: these optimisation problems do not quantify well the notion of predictability by an adversary.

One can instead, as in, e.g., [MRC+14], optimise over possible extensions $P_{ABE|XY}(abe|xy)$ of Alice and Bob's behaviour, where E is the random variable associable to Eve's output e, and maximise the probability of the event $E = A$ (local guessing probability) or $E = (A, B)$ (global guessing probability). Note that, when the eavesdropper has to to guess the output or output pair

associated to a specific input x or input pair (x, y), the ability to perform different measurements does not give an advantage, which is why we define the extension as $P_{ABE|XY}(abe|xy)$ and not $P_{ABE|XYZ}(abe|xyz)$. The global guessing probability problem then becomes:

$$\begin{aligned} G(AB|xy) = \max_{P_{ABE|XY}} \quad & \sum_{ab} P_{ABE|XY}(ab, e = (a, b)|xy), \\ \text{s.t.} \quad & \mathcal{I}(P_{AB|XY}) = I, \\ & P_{ABE|XY} \in \mathcal{Q}. \end{aligned} \tag{2.25}$$

where $P_{AB|XY} = \sum_e P_{ABE|XY}(abe|xy)$ is Alice-Bob marginal behaviour induced by the extension, and where the set $\mathcal{Q}$ is the straightforward extension of the set described in Def. 1 to tripartite behaviours. The no-signalling set (Def. 3), as well as the NPA sets $\mathcal{Q}_k$, can also be straightforwardly extended to tripartite behaviours. We can define similarly the local guessing probability.

Note that, using Baye's rule and the no-signalling condition between Eve and Alice-Bob, the following holds:

$$P(abe|xy) = P(e)P(ab|xye). \tag{2.26}$$

Setting $\alpha\beta = e$, and rewriting $P(ab|xy\alpha\beta)$ as $P^{\alpha\beta}(ab|xy)$, one obtains the following alternative definition for the guessing probability [BSS14]:

$$\begin{aligned} G(AB|xy) = \max_{\{P^{\alpha\beta}\}} \quad & \sum_{\alpha\beta} P^{\alpha\beta}(\alpha\beta|xy), \\ \text{s.t.} \quad & \mathcal{I}(\sum_{\alpha\beta} P^{\alpha\beta}) = I, \\ & \sum_{\alpha\beta ab} P^{\alpha\beta}(ab|xy) = 1, \\ & \forall\ \alpha, \beta,\ P^{\alpha\beta} \in \tilde{\mathcal{Q}}. \end{aligned} \tag{2.27}$$

where the weights $P(\alpha\beta)$ are absorbed into the behaviours $P^{\alpha\beta}$, which is why we impose that these behaviours now belong to the set of unnormalised quantum behaviours $\tilde{\mathcal{Q}}$ (or unnormalised NPA set $\tilde{\mathcal{Q}}_k$, or unnormalised no-signalling set $\tilde{\mathcal{NS}}$), and we add separately the normalisation constraint (third line). This makes the objective function of (2.27) linear.

The optimisation problems defined by Eqs. (2.25) and (2.27) are equivalent. The first one correspond to looking for the best extension of a behaviour, the

second one to the best decomposition. We will indifferently use both formulations in this thesis. As a function of I, they correspond to taking the concave hull of (2.20), and, as such, encompass well the notion of guessing probability. Note that, if (2.20) is already concave in I, the formulations (2.25) and (2.27) are superfluous.

In [NSPS14, BSS14], the authors noted that the constraint on the violation of a Bell inequality could be modified: one can impose that Eve's decomposition (or extension) yields a given marginal P^* for Alice and Bob, instead of imposing that this marginal yields a given Bell violation $\mathcal{I}(P^*) = I$. This operationally means that Alice and Bob check the full statistics of their black boxes, not only a linear functional of it. As the eavesdropper is more constrained, the guessing probability obtained in this manner is smaller: more randomness can be extracted. The optimisation problem then takes the form:

$$\begin{aligned} G(AB|xy) = \max_{\{P^{\alpha\beta}\}} \quad & \sum_{\alpha\beta} P^{\alpha\beta}(\alpha\beta|xy), \\ \text{s.t.} \quad & \sum_{\alpha\beta} P^{\alpha\beta} = P^*, \\ & \forall\ \alpha, \beta,\ P^{\alpha\beta} \in \tilde{\mathcal{Q}}. \end{aligned} \tag{2.28}$$

Note that adding the normalisation constraint is not necessary here, as it follows from the first constraint.

In order to give more information about (2.28), we now introduce a few basic notions of dual optimisation. Any SDP can be written in the following standard form:

$$\begin{aligned} & \max_X \operatorname{Tr}(CX), \\ & \text{s.t. } \forall i \in \{1, ..., m\},\ \operatorname{Tr}(A_i X) = b_i, \\ & \qquad X \succeq 0, \end{aligned} \tag{2.29}$$

where the optimisation variable is the symmetric matrix X, the objective function is characterised by the symmetric matrix C, and the constraints are defined by the symmetric matrices $A_1, ..., A_m$ and the real vector $b = (b_1, ..., b_m)$. Let's call (2.29) the primal problem. Then we define its associated dual problem as:

$$\begin{aligned} & \min_y b^\top y, \\ & \text{s.t. } \sum_{i=1}^m y_i A_i - C \succeq 0. \end{aligned} \tag{2.30}$$

These two problems are related by the weak duality theorem:

Theorem 3. *Let p be the optimal value of* (2.29) *and d be the optimal value of* (2.30)*. Then*

$$p \leq d. \tag{2.31}$$

Weak duality implies that solving an SDP can be achieved by finding a feasible point for the primal and a feasible point for the dual that achieve the same objective function value. A stronger theorem, called strong duality, states that, in some cases, the primal and the dual have the same solution. In particular:

Theorem 4. *If* (2.29) *is an linear program that admits a feasible point, then*

$$p = d. \tag{2.32}$$

Solving the dual problem of (2.28) provides the optimal Bell expression for certifying randomness [NSPS14, BSS14], i.e., the Bell expression such that (2.27) and (2.28) have the same value.

The guessing probability problem can be solved for several varying parameters, such as the underlying theory, the Bell expression and its associated value, or the observed behaviour. It can also be extended to a scenario with more than two observers (see Chapter 5). We now explain why this quantity is relevant, from both fundamental and applied perspectives.

2.3.2 Fundamental aspects of the guessing probability

Nonlocal theories are intrinsically random. This assertion can be certified, but also quantified, by the observation of a guessing probability strictly smaller than 1, even in the presence of an adversary with unlimited power. However, the relations between nonlocality and randomness are not trivial. Evaluating the guessing probability in various contexts enables us to shine a new light on a physical theory, now examined in terms of randomness. More precisely, it allows studying the relations of various nonlocal theories to one another, and of quantum physics amongst them. Moreover, within quantum physics, one can then classify various resources according to how good they are for certifying randomness.

Several theories imply the existence of nonlocal behaviours, as was illustrated by the examples of the quantum set $\mathcal{Q}$, the no-signalling set $\mathcal{NS}$, and the NPA relaxations of the quantum set $\mathcal{Q}_k$. Other theories can be constructed, and one way to relate them is to quantify their 'unpredictiveness' power. In [dlTHD$^+$15],

for instance, the authors show that maximally nonlocal theories do not permit maximal randomness, while quantum theory does. They then ask whether that singles out quantum theory amongst all other nonlocal theories, and answer in the negative.

Alternative models to the no-signalling and quantum bipartite dichotomic scenario can occur when one departs from the ideal situation where the two black boxes are characterised by a single bipartite dichotomic behaviour, and assume that each interaction with the devices is governed by a different distribution, that can depend on the previous interactions. This sequential approach leads to alternative sets of nonlocal behaviours. One can then compute guessing probabilities in such frameworks (see Chapter 3). Note that such sets are not only of fundamental interest, see Section 2.3.3 for details.

Within the framework of quantum theory, one can study how much randomness can be obtained from a given resource. The authors of [AMP12] proved that almost maximal global randomness can be certified from almost unentangled states. We address in Chapter 6 the question of whether maximal global randomness can be certified from any bipartite partially entangled qubit, i.e., any $|\Phi_\theta\rangle$.

One can connect various nonlocal theories and various quantum resources to the broad notion of randomness via the guessing probability. However, randomness is also valuable from a practical point of view, and how can certify randomness in a device-independent way via the guessing probability.

2.3.3 Link to randomness generation

Beyond its fundamental interest, randomness has a wide range of practical applications. Random numbers are useful for statistical sampling, video games or numerical simulations, but also for cryptography. In that case, it is paramount that these random numbers are private, i.e., that no information about them is known to a third party. We now explain how such a property can be certified in a device-independent way by quantum physics.

The ultimate goal of a randomness generation protocol is to produce a sequence of bits that is close to being uniformly distributed and uncorrelated to any information held by an external agent Eve. That typically involves several steps, and the terminology to describe them sometimes fluctuates. Let us now fix the terminology that we use in this thesis.

- **Randomness expansion** refers to the task of using an initial random bit string to generate a longer partially private random bit string. The quality

of the generated bit string is then quantified by its min-entropy.

- **Randomness extraction** or **privacy amplifcation** consists in taking a partially private and random bit string whose min-entropy is lower bounded, along with a random seed, and obtaining a shorter string that is (very close to) uniformly distributed and (almost) completely private, i.e., uncorrelated to any information held by an adversary.

- **Randomness amplification** aims to generate an (almost) uniform random bit from a partially random one (i.e., without an additional seed). This was proven to be impossible in the classical case [SV84], but can be achieved with quantum devices [CR12, GMdlT$^+$13].

The last task is somewhat distinct from the first two, and is not the subject of this thesis. Randomness expansion can be achieved in a device-independent way by using a Bell test [Col06, PAM$^+$10] and randomness extraction can be carried out with a randomness extractor [DPVR12]. These two tasks can then be articulated to obtain a bit string with the desired properties.

They both require some initial randomness. This observation might give the impression that device-independent randomness generation and certification is circular, and, as such, meaningless. That would be incorrect in two ways.

Firstly, the amount of randomness that is needed for generating the inputs of the Bell test and as the seed for the extractor is typically smaller than the one obtained as the final bit string. For that reason, one might study a randomness expansion protocol in terms of efficiency, i.e., comparing the randomness that we obtained to the randomness that we supplied.

Secondly, the nature of the initial randomness is different from the nature of the final randomness. Indeed, the randomness that is inputed can be public: it should not be correlated with the internal working of the quantum devices and of the extractor, but it can be known to an external agent. On the other hand, the value of the final random string is its privacy: an external agent has no information on this bit string, which can thus be used for cryptographic protocols. We take this approach in this thesis, which is why we talk about 'randomness generation', implying 'private randomness generation'.

The theory of extractors has been extensively studied. An extractor is defined with regards to a lower bound on the inputs' min-entropy m. In an adversarial scenario, the min-entropy of a random variable is the negative logarithm of the probability that the adversary correctly guesses the values of that random variable. This quantity thus depends on the nature of the correlations between

Alice-Bob and the adversary. An explicit extractor then gives a relation between m, the quality of the seed, and the length of the extracted string l.

In the case where Eve is classically correlated to Alice-Bob, the extractors that were designed without taking into account the (classical) side information E [BBC95, Tre01] are well suited [KR11]. If Eve shares quantum correlations with Alice-Bob, that is, if the side information E represents a quantum state, a novel theory of so-called quantum-proof extractors was developed [TSSR11, DPVR12]. Finally, if Eve, Alice and Bob are correlated in a supra-quantum way, privacy amplification by hashing was proven to be impossible if no additional constraints are imposed on the correlations [HRW13], but little is known if one makes the natural assumption that the correlations are time-ordered, i.e., that only past events can influence future events [AFTS12].

Randomness extractors are not within the scope of this thesis. Let us simply mention that one can achieve at least:

$$l = m + O(\log_2(1/\epsilon)), \tag{2.33}$$

for both classical side [ILL89] and quantum side [Ren05] information via two-universal hashing, and that one can obtain a longer extracted string if one quantifies the quality of the inputs via the ϵ-smooth version of the min-entropy [RW04], that consists in taking the maximal min-entropy in ball of size ϵ around the $\mathcal{ABE}$ correlations (in the case of classical side information) or around the $\mathcal{ABE}$ state (for quantum side information).

We now explain how one can relate the guessing probability problem and the min-entropy, in the case of classical side information. We then briefly mention the cases of quantum and supra-quantum side information.

A randomness generation protocol against an adversary with classical side information is depicted in Fig. 2.6. In that case, the conditional min-entropy is defined as:

$$H_{\min}(\mathbf{AB}|\mathbf{XY}, E) = -\log_2 \sum_{\mathbf{x},\mathbf{y},e} P_{\mathbf{XY}E}(\mathbf{x},\mathbf{y},e) \max_{\mathbf{a},\mathbf{b}} P_{\mathbf{AB}|\mathbf{XY}E}(\mathbf{ab}|\mathbf{xy}e). \tag{2.34}$$

Note that this quantity is also conditioned on the inputs on the Bell test. It corresponds to assuming that this information is accessible to the eavesdropper. It is thus the most conservative approach, as cryptographic protocols often requires that Alice and Bob reveal some (or all) of their inputs on a public channel.

A method to derive a lower bound on this n-round min-entropy based on the one-round guessing probabilities $G(AB|xy)$ was first presented in [PAM+10].

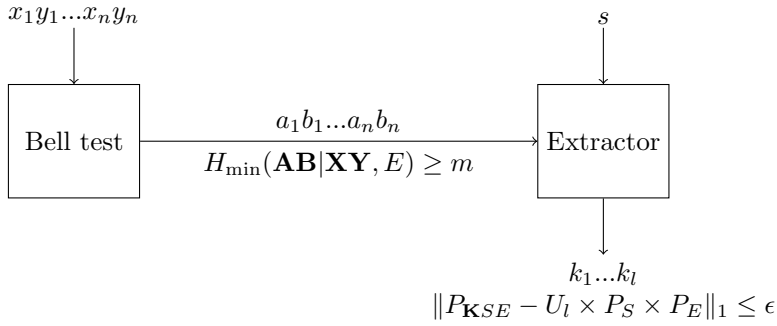


Figure 2.6: Schematic representation of a (private) randomness generation protocol with classical side information. U_l is the uniform distribution on $\{0,1\}^l$. The distance between the ideal distribution $U_l \times P_S \times P_E$ and the real one $P_{\mathbf{K}SE}$ is the variational distance. We take into account the seed s in the security requirement, which can be achieved with a so-called strong extractor.

It involves a statistical analysis on the n-round distribution via the Azuma-Hoeffding inequality, and allows to bound the min-entropy conditioned on the observation of a certain average value of the Bell inequality used in the guessing probability problem (2.25). Some mistakes in the derivation were then fixed in [PM13, FGS13].

The result essentially states that the n-round min-entropy is lower-bounded by n times $-\log_2[G(AB|XY)]$, where $G(AB|XY)$ is evaluated for a Bell violation slightly less nonlocal than the observed average, with some correction terms. Here, $G(AB|XY)$ denotes the worst case guessing probability for all inputs, i.e.:

$$G(AB|XY) = \max_{xy} G(AB|xy). \tag{2.35}$$

A generalisation of this bound was then derived in [NSBSP18]. In the same spirit as the generalisation from (2.27) to (2.28), it takes into account the possibility of bounding the min-entropy from a one-round randomness bound valid for the full Alice-Bob statistics, but also from a randomness bound valid for any number of Bell inequalities. Moreover, this randomness bound doesn't have to correspond to the worst case for all inputs, as in Eq. (2.35), but can be computed for any (more favorable) subset of inputs. The exact statement of this bound can be found in Chapter 4.

Assuming that the side information is classical is relevant in the case of

DIRNG, as DIRNG typically involves one user in one location: no quantum information needs to be sent over an insecure channel. In that case, the device-independent approach allows to certify randomness even if the devices are imperfect or not fully characterised, and even if an adversary takes advantage of these possible flaws or lack of knowledge to try and predict the outcomes of the Bell test [PM13]. However, in the case of DIQKD, where Alice and Bob exchange quantum information over a public channel, one has to assume that the eavesdropper can hold quantum side information.

In that case, that we do not consider in this thesis, a quantum version of the min-entropy was defined [Ren05]. It was proven to accurately quantify the number of bits that can be extracted against an adversary with quantum side information [KR11]. Moreover, it can also be interpreted as the negative logarithm of a quantum guessing probability [KRS09]. A way to lower bound this quantity was introduced in [AFDF$^+$18].

Finally, one can define a guessing probability problem similar to that of Eq. (2.28) for the whole sequence of Bell tests $G(\mathbf{AB}|\mathbf{xy})$. By definition of the min-entropy, one would thus have:

$$H_{\min}(\mathbf{AB}|\mathbf{xy}, E) = -\log_2[G(\mathbf{AB}|\mathbf{xy})] \tag{2.36}$$

Moreover, if Eve is equipped with a measurement choice Z, and can be correlated in a supra-quantum way with Alice-Bob:

$$H_{\min}(\mathbf{AB}|\mathbf{XY}, ZE) = -\log_2[\sum_{\mathbf{xy}} P_{\mathbf{XY}}(\mathbf{xy})G(\mathbf{AB}|\mathbf{xy})], \tag{2.37}$$

see Chapter 3 for details.

Chapter 3

Quantifying the randomness of copies of noisy Popescu-Rohrlich correlations

In a no-signalling world, the outputs of a nonlocal box cannot be completely predetermined, a feature that is exploited in nonlocality based quantum information protocols, such as DIRG or DIQKD. The relation between nonlocality and randomness can be formally quantified through the min-entropy, a measure of the unpredictability of the outputs that holds conditioned on the knowledge of any adversary that is limited only by the no-signalling principle. This quantity can easily be computed for the noisy PR-box, the paradigmatic example of nonlocality. In this Chapter, we consider the min-entropy associated to several copies of noisy PR-boxes. In the case where n noisy PR-boxes are implemented using n non-communicating pairs of devices, it is known that each PR-box behaves as an independent biased coin: the min-entropy per PR-box is constant with the number of copies. We show that this doesn't hold in more general scenarios where several noisy PR-boxes are implemented from a single pair of devices. In this case, the min-entropy per PR-box is smaller than the min-entropy of a single PR-box, and it decreases as the number of copies increases. The results of this Chapter are based on [BPA18b].

3.1 Introduction

Devices that are nonlocally correlated, i.e., which violate Bell inequalities, necessarily produce outcomes that cannot be perfectly determined [Val02]. This statement is true even according to theories that can deviate from the standard quantum formalism, provided that they satisfy the no-signalling principle according to which local measurements made on a subsystem cannot reveal information about measurements performed on distant subsystems.

This relation between nonlocality, randomness, and no-signalling can be illustrated through the paradigmatic example of the noisy PR-box. Suppose that Alice and Bob perform a Bell test with dichotomic inputs and outputs, governed by the behaviour:

$$\mathrm{PR}_v(ab|xy) = \begin{cases} 3/8 + v/8 & \text{if } a + b = xy \mod 2 \\ 1/8 - v/8 & \text{otherwise,} \end{cases} \tag{3.1}$$

parameterized by the number $v \in [-1, 1]$. The case $v = 1$ corresponds to the ideal PR-box, $v = -1$ to uniform white noise, and the intermediate cases to noisy-PR boxes given by a mixture of these two possibilities. The devices violate the CHSH inequality, hence are nonlocal, when $v \in]0, 1]$. They can be realized through measurement on a quantum state when $v \leq \sqrt{2} - 1$, with $v = \sqrt{2} - 1$ corresponding to Tsirelson-correlations, i.e., correlations reaching the maximal quantum violation of the CHSH inequality.

We can quantify how random Alice's outcome a is by considering how predictable it is to a third party Eve. Eve could hold information allowing her to guess Alice's outcome a with greater probability that what directly follows from the distribution (3.1). For instance, it could be that this distribution is realized as a mixture of underlying distributions which are individually less random than (3.1) and that Eve is aware of which one of these underlying distributions is currently realized. More generally, Eve could hold some physical system correlated to Alice's and Bob's devices and performing a measurement on her system could reveal useful information about Alice's outcome. Denoting z Eve's measurement choice and e the corresponding outcome, we can describe this situation through a tripartite distribution $P(abe|xyz)$, whose marginal distribution for Alice and Bob corresponds to the noisy PR-correlations: $\sum_e P(abe|xyz) = \mathrm{PR}_v(ab|xy)$.

It can easily be shown that, no matter what Eve's strategy is, the maximum

probability $G_1(A|x)[v]$ with which she can guess Alice's outcome a is[1]

$$G_1(A|x)[v] = 1 - \frac{v}{2}\,. \tag{3.2}$$

This value holds under the only assumption that Alice, Bob, and Eve's systems satisfy the no-signalling constraints

$$\begin{aligned} P(ab|xyz) &= P(ab|xy), \\ P(ae|xyz) &= P(ae|xz), \\ P(be|xyz) &= P(be|yz), \end{aligned} \tag{3.3}$$

stating that the input of one's party cannot affect the marginal distribution of the two other remote parties. Eq. (3.2) is proven in Appendix A.1 and Eve's optimal strategy is sketched in Fig. 3.1.

The optimal guessing probability (3.2) represents a measure of the randomness of noisy PR-correlations. It is strictly smaller than 1, and thus Alice's outcome cannot be perfectly predicted by Eve, when $v > 0$, i.e., when Alice's and Bob's devices are nonlocal. It is also common to use the min-entropy $H^{(1)}_{\min}(A|x,E)[v] = -\log_2 G_1(A|x)[v]$ to express the randomness of (3.2) in bits [KRS09]. For instance, the ideal PR-correlations have $H^{(1)}_{\min}(A|x,E)[1] = 1$ bit of randomness, while the Tsirelson-correlations have $H^{(1)}_{\min}(A|x,E)[\sqrt{2}-1] = 1-\log_2(3-\sqrt{2}) \simeq 0.335$ bits of randomness.

In this Chapter, we investigate the randomness of noisy-PR correlations in a scenario where Alice and Bob make n observations each, instead of a single one. This operationally corresponds to Alice and Bob using n times a single pair of devices, instead of a single one, either because they use n devices or a single device repeatedly n times. They thus end up with, respectively, input strings $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ and output strings $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$. We assume that Alice and Bob's observations are distributed according to

$$P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}_v(a_i b_i|x_i y_i)\,. \tag{3.4}$$

This means that, from Alice and Bob's perspective, their outputs are the same as if they had used n identical and independent copies of the noisy PR-correlations (3.1). This example was also studied in [FHSW10], where the authors investigate

[1] Anticipating a notation that we will use later on, the subscript "1" in $G_1(A|x)[v]$ refers to a single copy of the noisy PR-box (3.1).

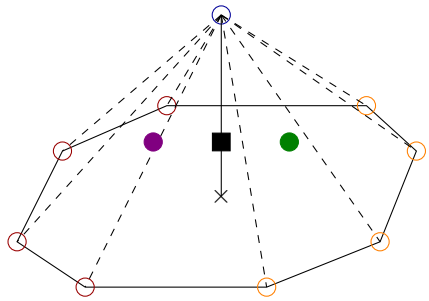


Figure 3.1: Schematic representation of the adversarial strategy that achieves the value given in Eq. (3.2). The base of the pyramid represents the CHSH facet of the local set. The eight extreme points on this facet are the eight deterministic strategies attaining CHSH=2. The blue point on top represents the PR-box. For some fixed inputs x, y, the local points on the left side (in red) yield the same value for a, say 0, and the ones on the right side (in orange) yield the other possible value, say 1. In order to guess the value of a, Eve can prepare either a mixture of the red and blue points (in purple), and guess $a = 0$, or a mixture of the orange and blue points (in green), and guess $a = 1$. On average, these two points reproduce Alice and Bob's expected distributions, PR_v, here depicted by a square.

its so-called local part. Note that, even though their results have some similarities with ours, there is no direct connection between the local part of some nonlocal correlation and its unpredictability.

We here ask how predictable Alice's string $\mathbf{a}$ is to some third party, Eve, under the sole assumption of no-signalling. In full generality, we can again characterise correlations among Alice, Bob, and Eve through a $2n + 1$-partite distribution $P(\mathbf{a}, \mathbf{b}, e|\mathbf{x}, \mathbf{y}, z)$, consisting of n input and output bits for Alice, n input and output bits for Bob, and a single input and output symbol for Eve.

There are, however, different ways to generalize the no-signalling conditions (3.3) to our $2n+1$-partite situation, depending on how Alice and Bob's experiment is performed (see Fig. 3.2). For instance, Alice and Bob could use n separated pairs of devices, where each pair $i = 1, \ldots, n$ receives inputs x_i, y_i and produces outputs a_i, b_i. They could use a single pair of devices n times in succession, where now x_i, y_i and a_i, b_i refers to the inputs and outputs at the ith round. A further possibility is that Alice holds some big device where she directly inputs n-bit strings $\mathbf{x}$ and get n-bits output strings $\mathbf{a}$, and similarly Bob holds a big device

accepting n-bit inputs $\mathbf{y}$ and producing n-bit outputs $\mathbf{b}$. To each such physical scenario is associated a different set of no-signalling constraints corresponding to limitations on how the input x_i (or y_i) can causally influence the output strings $\mathbf{a}$ and $\mathbf{b}$. In what follows, we will define in more details four natural scenarios and their associated no-signalling constraints.

In all cases, one possible strategy for Eve is to guess each of Alice's output a_i independently using the optimal single-copy strategy yielding (3.2). However, there may exist clever strategies that perform better than this independent guessing strategy. This is so even though the correlations (3.4) look identical and independent from Alice's and Bob's perspective, because they need not look that way from Eve's point of view. Indeed, the probabilities $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}, e, z)$ conditioned on Eve's knowledge do not need to take a product form, only their average $\sum_e P(e|z)P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}, e, z) = P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$, corresponding to tracing out Eve, should. In particular, Eve can design the correlations $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}, e, z)$ in such a way that the distribution of an output pair (a_i, b_i) is correlated with other values of inputs and outputs. This enables Eve to increase the predictability of some particular sequences, conditioned on the value of e, while keeping Alice and Bob's marginal distributions unchanged.

We show that this is indeed what happens for several no-signalling scenarios of interest. The single-copy guessing probability (3.2) thus does not correctly reflect the randomness of noisy PR-boxes in a situation involving n copies of such correlations.

Beside its fundamental interest, this investigation is also motivated by the problem of understanding better the security of quantum key distribution and quantum random number generation against no-signalling adversaries, whose status is not clear at the moment [AFTS12, SW16]. Previous works have looked at how much information a no-signalling adversary can obtain about the outcomes of n PR-boxes after privacy amplification [AFHTS12]. We look here at her information before privacy amplification, i.e., on the raw output string. Though the results that we present do not have yet direct implications for the security of quantum key distribution and quantum random number generation schemes, they contribute to a better characterisation of adversarial strategies.

3.2 Definitions

Before we present and discuss our results, we introduce here the problem that we consider in more details.

3.2.1 General scenario

We use subscripts to denote certain sub-strings of n-bit strings, e.g. $\mathbf{x}_{\leq i} = (x_1, \ldots, x_i)$, $\mathbf{x}_{>i} = (x_{i+1}, \ldots, x_n)$ or $\mathbf{x}_{\setminus i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. The subscript 0 corresponds to the empty string: $\mathbf{x}_0 = \emptyset$.

Alice, after interacting n times with one or several devices, ends up with input and output strings $\mathbf{x}$ and $\mathbf{a}$. Similarly, Bob ends up with input and output strings $\mathbf{y}$ and $\mathbf{b}$. We assume, as in Eq. (3.4), that the joint probabilities $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ correspond to n-copies of noisy PR-correlations.

We assume that Eve holds a system that may be correlated to Alice's and Bob's devices, a situation that can be described, as in the introduction, through a distribution $P(\mathbf{a}, \mathbf{b}, e|\mathbf{x}, \mathbf{y}, z)$ that is compatible with Alice and Bob marginals. Under the assumption that these correlations cannot be used for signalling between Eve and Alice-Bob, we can describe things in an alternative, convenient way that does not directly involves Eve's input z. Indeed, as explained in [ACP$^+$16], any measurement that Eve can perform on her system can be interpreted as a choice of a convex decomposition

$$\sum_e P(e) P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}_v(a_i b_i | x_i y_i) \tag{3.5}$$

of Alice's and Bob's devices and her measurement outcome e can be interpreted as indicating one part of this decomposition. Conversely, any convex decomposition (3.5) of Alice and Bob's system can be realized by Eve by choosing an appropriate measurement on her system. From now on, we adopt this view.

The components $P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ in the above decomposition are not arbitrary but should satisfy certain no-signalling constraints reflecting the causal relations that follow from the way Alice and Bob use their devices. We consider four types of such no-signalling constraints.

Definition 5 (Full-NS)**.** *The probabilities $P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ are fully no-signalling (Full-NS) if, for every $1 \leq i \leq n$,*

$$P^e(\mathbf{a}_{\setminus i}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}_{\setminus i}, \mathbf{b}|\mathbf{x}_{\setminus i}, \mathbf{y}), \tag{3.6}$$

$$P^e(\mathbf{a}, \mathbf{b}_{\setminus i}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}, \mathbf{b}_{\setminus i}|\mathbf{x}, \mathbf{y}_{\setminus i})\,. \tag{3.7}$$

In the above definition, it is to be understood that Eq.(3.6) holds for all possible values of $\mathbf{a}_{\setminus i}, \mathbf{b}, \mathbf{x}, \mathbf{y}, e$ and Eq.(3.7) for all possible values of $\mathbf{a}, \mathbf{b}_{\setminus i}, \mathbf{x}, \mathbf{y}, e$.

The marginal distribution $P^e(\mathbf{a}_{\setminus i}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ is obtained by summing the whole probability table of Alice and Bob over the missing variables: $P^e(\mathbf{a}_{\setminus i}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{a}_i} P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ and analogously for $P^e(\mathbf{a}, \mathbf{b}_{\setminus i}|\mathbf{x}, \mathbf{y})$. The other definitions that we introduce below should be understood similarly.

This condition corresponds to having $2n$ parties which satisfy all possible pairwise no-signalling conditions. It is operationally equivalent to using $2n$ boxes that are all causally independent, i.e., no communication is allowed between any of them, even though they can be correlated [MPA11, HRW10]. See Fig. 3.2 (a) for a schematic representation of this scenario.

Definition 6 (ABNS)**.** *The probabilities $P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ are Alice-Bob no-signalling (ABNS) if*

$$P^e(\mathbf{b}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{b}|\mathbf{y}), \tag{3.8}$$
$$P^e(\mathbf{a}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}|\mathbf{x})\,. \tag{3.9}$$

In this case, no-signalling holds only between Alice and Bob, i.e., there is no communication between them. It means that the inputs used by Bob cannot be inferred from Alice's marginal distribution, even if the information from all the rounds is grouped together, and vice-versa. However, there is no constraint on the internal structure of Alice's or Bob's own marginal. For instance, output a_1 could depend on the values of all the inputs $\mathbf{x} = (x_1 \dots x_n)$.

It is equivalent to considering one big device on Alice's side (respectively Bob's side), that receives as input the string $\mathbf{x} = (x_1 \dots x_n)$ (resp. $\mathbf{y}$) and produces at once the output string $\mathbf{a} = (a_1 \dots a_n)$ (resp. $\mathbf{b}$), or n devices on each side that are used in parallel and can communicate freely amongst themselves [HRW13]. This condition is schematically depicted in Fig. 3.2 (b).

Definition 7 (TONS)**.** *The probabilities $P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ are time-ordered no-signalling (TONS) if, for every $0 \leq i < n$*

$$P^e(\mathbf{a}_{\leq i}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}_{\leq i}, \mathbf{b}|\mathbf{x}_{\leq i}, \mathbf{y}), \tag{3.10}$$
$$P^e(\mathbf{a}, \mathbf{b}_{\leq i}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}, \mathbf{b}_{\leq i}|\mathbf{x}, \mathbf{y}_{\leq i})\,. \tag{3.11}$$

In this case, no-signalling holds between Alice and Bob as for ABNS (take $i = 0$). In addition, future rounds (which corresponds to values greater than i) have no influence on past rounds (which corresponds to values smaller than i) on each

side. It describes the situation where two devices are separated from each other during the entire run of the experiment and are used sequentially, while keeping a memory of the past events [AFTS12, SW16]. The schematic representation of this condition can be found in Fig. 3.2 (c).

Note that Full-NS $\subset$ TONS $\subset$ ABNS.

Definition 8 (WTONS)**.** *The probabilities* $P^e(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ *are weakly time-ordered no-signalling (WTONS) if for all* $0 \leq i < n$*,*

$$P^e(\mathbf{a}_{\leq i}, \mathbf{b}_{\leq i+1}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}_{\leq i}, \mathbf{b}_{\leq i+1}|\mathbf{x}_{\leq i}, \mathbf{y}_{\leq i+1}), \tag{3.12}$$

$$P^e(\mathbf{a}_{\leq i+1}, \mathbf{b}_{\leq i}|\mathbf{x}, \mathbf{y}) = P^e(\mathbf{a}_{\leq i+1}, \mathbf{b}_{\leq i}|\mathbf{x}_{\leq i+1}, \mathbf{y}_{\leq i})\,. \tag{3.13}$$

This condition is a weakened version of the time-ordered-no-signalling condition, i.e. TONS $\subset$ WTONS. Future rounds cannot influence past round, and no-signalling holds at each individual round, but, contrarily to ABNS and TONS, no-signalling between Alice and Bob does not hold throughout the entire run of the experiment. It means that Alice's marginal at round i is independent of $\mathbf{x}_{>i}$ and $\mathbf{y}_{\geq i}$, but can depend on $\mathbf{x}_{\leq i}$ and $\mathbf{y}_{<i}$, and likewise for Bob. It describes the situation where two devices are used sequentially and have memory, and where these two devices can moreover communicate between successive rounds [AFHTS12]. See Fig. 3.2 (d) for a schematic representation.

The TONS condition naturally emerges if the two devices can be shielded from each other during the entire experiment, e.g., if n pairs of entangled particles are stored in memory. Yet in many practical situations, pairs of entangled particles are produced one round after the other and distributed to each device. This requires that the devices be opened between each round, at which point some communication between the two devices could happen. WTONS characterises this situation.

Note that if we consider, as in the WTONS scenario, that communication between the boxes cannot be prevented between the successive rounds, one could also argue that one could not prevent the outcome bits from directly leaking to Eve, thus rendering the notion of guessing probability irrelevant. This point is pertinent in the case of protocols such as DIQKD, where Alice and Bob are indeed two distant agents aiming to share some private bits at distant locations. In this case, there is indeed no reason to believe that the information flowing from Alice to Bob could not also flow from Alice to Eve. However, for protocols such as DIRNG, Alice and Bob can be thought of as two fictional agents in a single laboratory, as the goal is here to obtain private bits in a unique location.

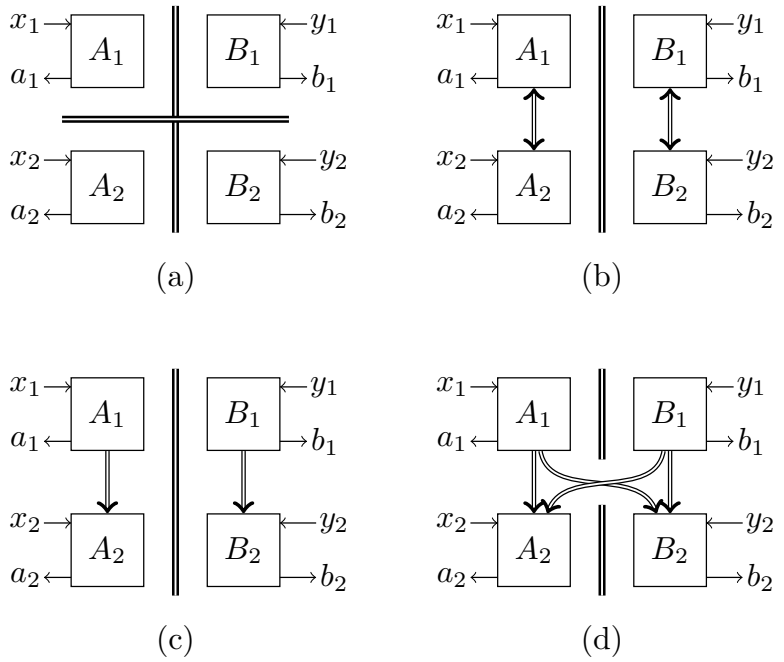


Figure 3.2: Representations of the Full-NS (a), ABNS (b), TONS (c) and WTONS (d) conditions. The double lines indicate that there is no information flow between the corresponding boxes, while the double arrows indicate the direction of the information flow.

In this situation, we believe that the TONS and WTONS scenarios are two relevant models.

3.2.2 Quantifying randomness

We quantify the randomness of Alice's output string **a** via the local form of the guessing probability problem defined by Eq.(2.28), that is, when the complete Alice-Bob behaviour is fixed. Eve's optimal guessing probability is thus the solution of the following optimisation problem:

$$G_n(\mathbf{A}|\mathbf{x}^*,\mathbf{y}^*)[v] = \max_{\{P(\boldsymbol{\alpha}),P^{\boldsymbol{\alpha}}\}} \sum_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\mathbf{a}=\boldsymbol{\alpha}|\mathbf{x}^*,\mathbf{y}^*) \tag{3.14}$$

$$\text{s.t.} \quad \sum_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}_v(a_i,b_i|x_i,y_i)$$

$$\forall \alpha,\ P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) \text{ is NS}$$

where NS denotes one of the no-signalling constraints NS={Full-NS, ABNS, TONS, WTONS}, depending on which scenario is considered. Note that here, we don't absorb the weights $P(\boldsymbol{\alpha})$ into the behaviours $P^{\boldsymbol{\alpha}}$, because they can be dealt with in a simple way, see Appendix A.4 for details.

It is implicit in the above formulation that Eve's choice of convex decomposition – and thus that the optimal guessing probability – depends on the inputs $\mathbf{x}^*$ and $\mathbf{y}^*$ that are chosen by Alice and Bob. We therefore assume that the specific inputs $\mathbf{x}^*$ and $\mathbf{y}^*$ used by Alice and Bob are communicated to Eve. Indeed, our aim is to quantify the fundamental, intrinsic randomness generated at Alice's side, even in a situation where all details of the experimental set-up are known to Eve. From an applied point of view, it also means that this quantity is relevant for a protocol where some actions are taken based on some specific values of inputs $(\mathbf{x}^*,\mathbf{y}^*)$, fixed in advance: the bound on the predictability is valid even if the protocol is known to Eve.

The optimal guessing probability $G_n(\mathbf{A}|\mathbf{x}^*,\mathbf{y}^*)[v]$ may therefore depend on the input choices $\mathbf{x}^*$ and $\mathbf{y}^*$ and there could thus be different possible ways to quantify the randomness of Alice's output: e.g., by considering the worst-case over all inputs choices or the expected guessing probability with respect to some probability distribution for Alice and Bob's inputs. In our case, however, thanks to the symmetries of the noisy PR-correlations (3.1), the same optimal value $G_n(\mathbf{A}|\mathbf{x}^*,\mathbf{y}^*)[v]$ is obtained for any possible choice of inputs $\mathbf{x}^*$ and $\mathbf{y}^*$. Indeed, as we show in Appendix A.2, given any solution to (3.14) for a given pair of inputs $\mathbf{x}^*,\mathbf{y}^*$, one can construct a corresponding solution for any other pairs of inputs that yields the same guessing probability. Thus we can simply quantify the randomness of Alice's output through the guessing probability associated to any given input choices. For specificity, we will use the choice $\mathbf{x}^* = \mathbf{y}^* = \mathbf{0} = (0_1,\ldots,0_n)$ in the following, and for simplicity, we will write $G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v] = G_n(v)$.

Note that, even if in our problem the optimal guessing probability for any input choices of Alice and Bob $G_n(v)$ is the same, the particular convex decom-

position achieving this optimal value will vary with the choice of inputs. If Eve is equipped with a measurement choice Z, she can remotely choose the optimal decomposition by selecting a measurement on her system when she is informed about Alice and Bob's input choices. This gives rise to a tripartite behaviour $P_{\mathbf{AB}E|\mathbf{XY}Z}$. Writing $\{P^{\alpha,\mathbf{x}^*,\mathbf{y}^*}\}_\alpha$ the argument of the maximum $G(\mathbf{A}|\mathbf{x}^*,\mathbf{y}^*)$, $P_{\mathbf{AB}E|\mathbf{XY}Z}$ is such that:

$$P_{\mathbf{AB}|\mathbf{XY},Z=(\mathbf{x}^*,\mathbf{y}^*),E=\alpha} = P^{\alpha,\mathbf{x}^*,\mathbf{y}^*}. \tag{3.15}$$

Moreover, $P_{\mathbf{AB}E|\mathbf{XY}Z}$ obeys no-signalling between Alice and Bob because $P^{\alpha,\mathbf{x}^*,\mathbf{y}^*}$ does, and between Alice-Bob and Eve because of the first constraint of (3.14). The conditional min-entropy:

$$H_{\min}^{(n)}(\mathbf{A}|\mathbf{X},\mathbf{Y},Z,E) = -\log_2[\sum_{\mathbf{x}^*\mathbf{y}^*ze} P_{\mathbf{XY}ZE}(\mathbf{x}^*\mathbf{y}^*ze)\max_{\mathbf{a}} P_{\mathbf{A}|\mathbf{XY}ZE}(\mathbf{a}|\mathbf{x}^*\mathbf{y}^*ze)] \tag{3.16}$$

for that distribution is thus:

$$H_{\min}^{(n)}(\mathbf{A}|\mathbf{X},\mathbf{Y},Z,E) = -\log_2[\sum_{\mathbf{x}^*\mathbf{y}^*} P_{\mathbf{XY}}(\mathbf{xy})\sum_{ze} P_{ZE}(ze)G(\mathbf{A}|\mathbf{x}^*\mathbf{y}^*)] \tag{3.17}$$

$$= -\log_2[\sum_{\mathbf{x}^*\mathbf{y}^*} P_{\mathbf{XY}}(\mathbf{xy})G(\mathbf{A}|\mathbf{x}^*\mathbf{y}^*)], \tag{3.18}$$

by normalisation, which is equal to

$$H_{\min}^{(n)}(\mathbf{A}|\mathbf{X},\mathbf{Y},Z,E) = -\log_2[G_n(v)] \tag{3.19}$$

in our case, because the guessing probabilities are the same for all $(\mathbf{x}^*,\mathbf{y}^*)$.

Thus, $G_n(v)$ correctly reflects the probability with which Eve can guess Alice's output in the most general scenario, and we now write

$$H_{\min}^{(n)}(v) = -\log_2[G_n(v)]. \tag{3.20}$$

However, this requires Eve to hold some "coherent memory", and to delay her measurement until when she is informed about Alice's and Bob's inputs. One could also consider, as in [PMLA13], a situation where Eve has no such "coherent memory" and is forced to commit to a decomposition before Alice's and Bob's inputs are known. Here we choose to quantify randomness in the former scenario because it corresponds to the worst possible setting where Eve's knowledge is maximal. Furthermore, it also corresponds to the scenario where

the security of RNG and QKD against no-signalling adversaries is not clearly established.

Finally, note that in the case of the Full-NS, ABNS, and TONS constraints, $P^{\boldsymbol{\alpha}}(\mathbf{a}|\mathbf{x}^*,\mathbf{y}^*) = P^{\boldsymbol{\alpha}}(\mathbf{a}|\mathbf{x}^*)$ and thus Eve's strategy does not actually need to depend on the knowledge of Bob's input $\mathbf{y}^*$. This, however, is not necessarily the case for the WTONS constraints for which no-signalling does not hold between Alice and Bob. This is why we include explicitly $\mathbf{y}^*$ in (3.14).

3.3 Basic observations and known results

Before presenting our actual results – the optimal solutions to (3.14) for different values of n, noise levels v, and different no-signalling conditions – let us make some basic observations.

3.3.1 Bounds on $G_n(v)$ from $G_1(v)$

For $n = 1$, all the no-signalling conditions NS={Full-NS, ABNS, TONS, WTONS} that we have introduced reduce to the usual no-signalling conditions between Alice and Bob:

$$P^e(a_1|x_1,y_1) = P^e(a_1|x_1) \tag{3.21}$$
$$P^e(b_1|x_1,y_1) = P^e(b_1|y_1)\,. \tag{3.22}$$

As we have claimed in the introduction, the optimal guessing probability $G_1(v)$ is known in this case and is given by Eq. (3.2).

Before attempting to find the guessing probabilities $G_n(v)$ for values of $n > 1$, we can already observe that they necessarily satisfy the trivial bounds

$$G_1^n(v) \leq G_n(v) \leq G_1(v) \tag{3.23}$$

or explicitly

$$\left(1-\frac{v}{2}\right)^n \leq G_n(v) \leq 1-\frac{v}{2}\,. \tag{3.24}$$

The lower-bound $G^n(v) \geq G_1^n(v)$ follows from the fact that a possible strategy is for Eve to guess each output bit of Alice a_i independently using the optimal strategy for a single copy of PR-correlations. The probability to guess correctly the entire string $\mathbf{a} = (a_1,\ldots,a_n)$ is then simply the product of the probability to guess correctly each bit independently. There could be, however, more clever

strategies, hence this only represents a lower-bound on the n-copy guessing probability $G_n(v)$.

The upper-bound $G_n(v) \leq G_1(v)$ follows from the fact that the probability to guess correctly the entire n-bit string $\mathbf{a}$ should not be higher than the probability to guess only one of the a_i.

For $v = 0$, corresponding to the point at which the noisy PR-correlations become local, the lower-bound and upper-bound coincide and give the trivial value $G_n(0) = 1$, as expected since any local correlations admit a purely deterministic explanation.

For $v = 1$, corresponding to perfect PR-correlations, it is possible to show that the lower-bound is saturated, i.e., $G_n(1) = (1/2)^n$. This follows from the fact that the product of n perfect PR-correlations is a vertex of the polytopes associated with any of the no-signalling constraints NS={Full-NS, ABNS, TONS, WTONS}, see Appendix A.3.

The values $G_n(v)$ for the different no-signalling constraints that we consider here thus all coincide at the extremities of the interval $v \in [0, 1]$ and our problem is to understand how the guessing probability varies as a function of n for $0 < v < 1$.

3.3.2 $G_n(v)$ in the Full-NS scenario

For the Full-NS scenario, it happens that the independent strategy discussed above is actually the optimal strategy. This directly follows from the results of Appendix A of [MRC+14], where it is shown that for every $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$ that is Full-NS, the following bound holds

$$G_n(v) \leq \sum_{\mathbf{a},\mathbf{b},\mathbf{x},\mathbf{y}} \prod_{i=1}^{n} \beta(a_i, b_i, x_i, y_i) P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}) \,, \tag{3.25}$$

where the coefficients β are defined as

$$\beta(a, b, x, y) = \begin{cases} 1/8 & \text{if } a + b = xy \mod 2 \\ 5/8 & \text{otherwise.} \end{cases} \tag{3.26}$$

In the case where $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}(a_i, b_i|x_i, y_i)$, it is easily seen that this yields $G_n(v) \leq \left(1 - \frac{v}{2}\right)^n$. Since this value can be trivially attained with the independent strategy discussed above, we have that

$$G_n(v) = \left(1 - \frac{v}{2}\right)^n \,. \tag{3.27}$$

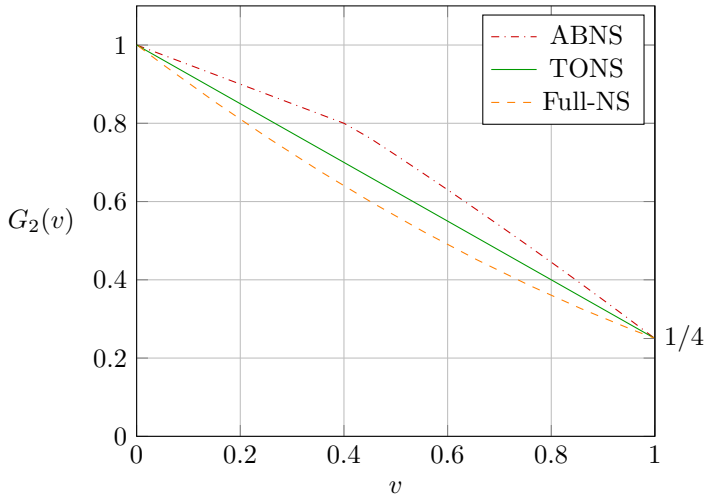


Figure 3.3: Guessing probabilities for $n = 2$, for the ABNS, TONS and Full-NS scenarios. The guessing probability for the WTONS conditions is the same as for the TONS conditions.

The min-entropy

$$H_{\min}^{(n)}(v) = -\log_2 G_n(v) = -n \log_2 \left(1 - \frac{v}{2}\right) = n H_{\min}^{(1)}$$

thus scales linearly with n: each new use of the noisy-PR correlations brings $H_{\min}^{(1)}$ new bits of randomness. Interestingly, we show below that this is no longer the case in the other no-signalling scenarios that we consider.

3.4 Results

The optimisation problem (3.14) is a linear program. This is easily seen, as explained in Section 2.3.1, by rewriting it in term of the unnormalized probabilities $\tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = P(\boldsymbol{\alpha}) P^{\boldsymbol{\alpha}}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$. For $n = \{2, 3, 4, 5\}$, we numerically solved this linear program for the three sets ABNS, TONS, WTONS.

We find in each case that the optimal guessing probability is higher than the one obtained with the independent strategy corresponding to the lower-bound in (3.24). Furthermore, for the cases $n = \{2, 3\}$, we solve (3.14) by finding explicit

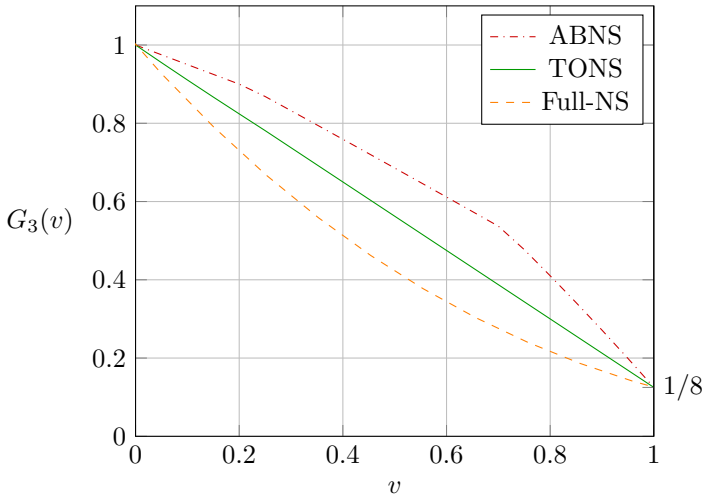


Figure 3.4: Guessing probabilities for $n = 3$. The change of behaviour at $v = -2 + \sqrt{5}$ indicated in Table 3.1 for the TONS and WTONS scenarios is not apparent because the polynomial is very close to the line in this region.

solutions to its primal and dual forms, and thus obtain the analytical expressions of $G_n(v)$.

3.4.1 Solutions for $n = \{2, 3\}$

The analytical solutions to the optimisation problem (3.14) are given in Table 3.1, and are plotted as a function of v in Figures 3.3 and 3.4. For the Full-NS scenario, we recover, as expected, the value (3.27) corresponding to the independent strategy. In the three other cases ABNS, TONS, WTONS, we find that the guessing probability is strictly higher than this value for all $0 < v < 1$.

The solutions are detailed in [BPA18a]. We now make a few observations. First of all, for $n = 2$ and $v \leq \sqrt{2} - 1$, the guessing probability $G_2(v)$ in the ABNS scenario saturates the trivial upper-bound in (3.24) given by the single-round guessing probability, i.e., $G_2(v) = G_1(v) = 1 - v/2$. This establishes, independently of our dual solutions, that our explicit strategy is optimal in this case.

Conceptually, it is surprising that the guessing probability does not decrease from $n = 1$ to $n = 2$ as it means that it is not more difficult for Eve to guess

$n = 2$, ABNS:

$$G_2(v) = \begin{cases} 1 - \frac{1}{2}v & \text{if } v \leq \sqrt{2} - 1 \\ \frac{9}{8} - \frac{3}{4}v - \frac{1}{8}v^2 & \text{if } v \geq \sqrt{2} - 1 \end{cases}$$

$n = 2$, TONS and WTONS:

$$G_2(v) = 1 - \tfrac{3}{4}v$$

$n = 3$, ABNS:

$$G_3(v) = \begin{cases} 1 - \frac{1}{2}v & \text{if } v \leq v_1 \\ \frac{67}{64} - \frac{45}{64}v - \frac{3}{64}v^2 + \frac{1}{64}v^3 & \text{if } v_1 \leq v \leq v_2 \\ \frac{41}{32} - \frac{27}{32}v - \frac{9}{32}v^2 - \frac{1}{32}v^3 & \text{if } v \geq v_2 \end{cases}$$

where v_1 is the unique root of $x^3 - 3x^2 - 13x + 3$ in $[0, 1]$ ($v_1 \approx 0.22038$) and v_2 the unique root of $x^3 + 5x^2 + 3x - 5$ in $[0, 1]$ ($v_2 \approx 0.70928$).

$n = 3$, TONS and WTONS:

$$G_3(v) = \begin{cases} 1 - \frac{29}{32}v + \frac{1}{8}v^2 + \frac{1}{32}v^3 & \text{if } v \leq \sqrt{5} - 2 \\ 1 - \frac{7}{8}v & \text{if } v \geq \sqrt{5} - 2 \end{cases}$$

Table 3.1: Analytical values of the guessing probabilities for $n = 2, 3$.

two outcome bits of Alice than it is to guess a single one. More surprisingly, the region where this happens corresponds to $v \leq \sqrt{2} - 1$, i.e., to the region where the noisy PR-correlations admit a quantum representation. We do not know whether this is merely a coincidence or whether it has some deeper meaning about the structure of the quantum set.

For $n = 3$, there is again a region, corresponding to $v \leq 0.22038$, where $G_3(v) = G_1(v)$. This region is smaller than the previous one, but on the other hand, Eve can now guess three successive bits of Alice with the same error probability as when guessing a single one.

For the TONS and WTONS scenarios, we find that the two solutions coincide. Interestingly, we find that the optimal solution in the case $n = 2$ is linear in v, as for $n = 1$. For $n = 3$, this is only true if v is above the threshold $v \geq \sqrt{5} - 2$.

We now intuitively explain how the strategies we have found work and the origin of this linear behaviour.

In our model, Eve distributes the correlations for Alice and Bob and can adapt the decomposition for each round depending on what happened in the previous rounds. For the first round, there is no past, so she prepares the mixture of extremal local and nonlocal points compatible with Alice and Bob's probabilities depicted in Figure 3.1.

The distribution for the second round depends on what happened in the first one [2]. If Alice's first output is such that Eve's guess is correct, the devices on the second round behave in a more predictable way, i.e., their correlations correspond to a more local point. This allows Eve to improve her guess on the two generated outputs. On the other hand, if Alice's first output is such that Eve's guess is wrong, the subsequent events are of no importance to the value of the guessing probability: the devices can be maximally nonlocal, i.e., a PR-box.

These different possibilities can then combine in such a way that Alice and Bob's marginal distributions are as expected, if Eve accurately adjusts the amount of nonlocality in the second round based on the value of v. For $n = 2$, the balance is such that the guessing probability is linear in v.

One could hope to straightforwardly extend this strategy to any number of rounds and that it would imply that the guessing probability be equal to $1 - (2^n - 1)/(2^n) \cdot v$ for all n. This is however not the case. To understand why, note that, in order to constantly improve her guess, Eve needs to prepare distributions that have more and more predictable outcomes, i.e., points that are closer and closer to the local set. But when a point is local, its outcomes are perfectly known to Eve: its predictability cannot increase anymore. We observe that, when this happens at some round, $G_n(v)$ is less than $1-(2^n-1)/(2^n)\cdot v$ for subsequent rounds. This phenomenon happens after a certain number of rounds, which depends on the value of v. For $n = 3$, we observe it for $v \leq -2 + \sqrt{5}$.

3.4.2 Solutions for $n = \{4, 5\}$

We then numerically solved (3.14) for $n = \{4, 5\}$. The results are plotted in Figures 3.5 and 3.6. In this case, we did not attempt to find the analytical expressions of $G_4(v)$ and $G_5(v)$: keeping track of the dual's variables, which grow exponentially with n, becomes demanding, while a numerical result is sufficient for our purpose.

[2] Let us stress that Eve doesn't need to acquire this knowledge for the strategy to be valid. This is merely a way to give an intuition about the strategy by decomposing it sequentially, while the attack is entirely designed prior to the experiment.

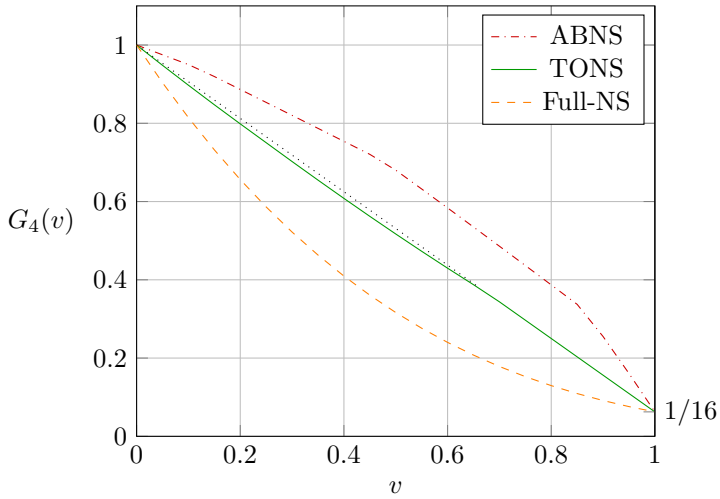


Figure 3.5: Guessing probabilities for $n = 4$. We add the line interpolating $(0, 1)$ and $(1, \frac{1}{16})$ (black dotted line) to emphasize the breakdown of linear dependence of the TONS guessing probability for some v.

As before, we observe that for v small enough, there is a region, that gets smaller as n increases, where $G_5(v) = G_4(v) = G_1(v)$ in the ABNS scenario.

For the TONS and WTONS scenarios, the guessing probability depends linearly on v (as $1 - 15/16 \cdot v$ for $n = 4$ and as $1 - 31/32 \cdot v$ for $n = 5$) when v is large enough. The minimal v for which this happens increases and gets closer to 1 as n increases.

However, while for $n \leq 3$, the guessing probability is the same for the TONS and WTONS scenarios, this is no longer the case when $n \geq 4$, except in the linear regime for v close to 1. The difference between the TONS and WTONS values is not visible on the graphs, which is why we highlight it in the following tables:

	v	0.05	0.1	0.15	0.2
$G_4(v)$	$WTONS$	0.9487	0.8981	0.8482	0.7990
	$TONS$	0.9481	0.8972	0.8473	0.7985

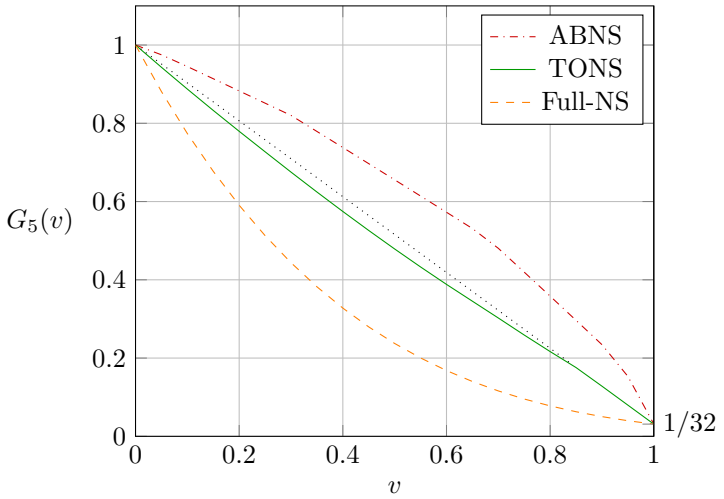


Figure 3.6: Guessing probabilities for $n = 5$. We add the line interpolating $(0, 1)$ and $(1, \frac{1}{32})$ (black dotted line) to emphasize the breakdown of linear dependence of the TONS guessing probability for some v.

	v	0.05	0.1	0.15	0.2
$G_5(v)$	$WTONS$	0.9451	0.8913	0.8387	0.7865
	$TONS$	0.9431	0.8874	0.8328	0.7795

Carrying out the numerical optimisation for larger n becomes computationally too demanding, as the number of variables and constraints grows exponentially with n. However, the results obtained for small n already have implications for all n, as explained below.

3.4.3 Implications for all n

For the ABNS, TONS, and WTONS scenarios, we have found in the previous subsections that, contrarily to what happens in the Full-NS scenario, the independent strategy is not the optimal strategy for $n = \{2, 3, 4, 5\}$, i.e., $G_n(v) > G_1^n(v)$.

This implies in particular that one can improve the lower-bound $G_n(v) \geq G_1^n(v)$ for all n, as, instead of considering strategies where Eve guesses independ-

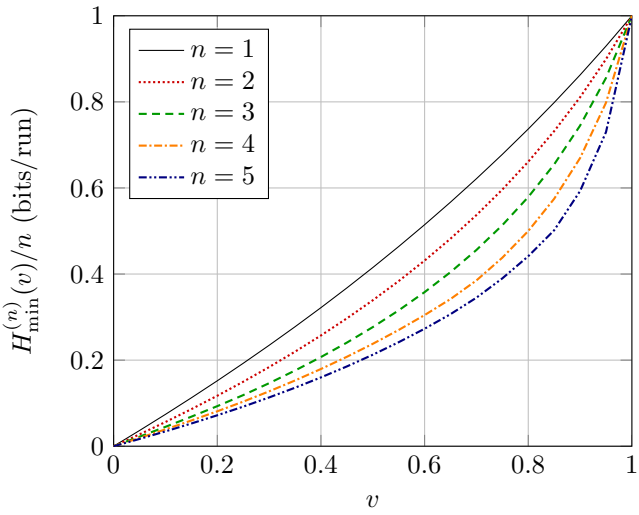


Figure 3.7: Min-entropy rates obtained when Eve is able to guess individual outcomes bits, and pairs, triples, quadruples and quintuples of outcome bits, in the TONS scenario. The curves for the WTONS scenario are virtually the same, as the guessing probabilities for WTONS are either the same or very close to the ones for TONS.

ently each individual outcome bit of Alice, one can now consider strategies where Eve guesses independently pairs, triples, quadruples or quintuples of outcome bits of Alice. For instance if $n = 5k$, Eve can guess every successive quintuple of outcomes independently, and we have thus the lower-bound

$$G_n(v) = G_{5k}(v) \geq G_5^k(v)\,. \tag{3.28}$$

In terms of the min-entropy per run this corresponds to the lower-bound

$$\frac{H_{\min}^{(n)}(v)}{n} = \frac{H_{\min}^{(5k)}(v)}{5k} \leq \frac{H_{\min}^{(5)}(v)}{5}\,, \tag{3.29}$$

which is strictly smaller than the single-run min-entropy: $H_{\min}^{(n)}(v)/n < H_{\min}^{(1)}(v)$, as illustrated in Figures 3.7 and 3.8.

In other words, for multiple uses of the noisy PR-correlations, each instance of the PR-correlations carry less entropy than what one would have naively

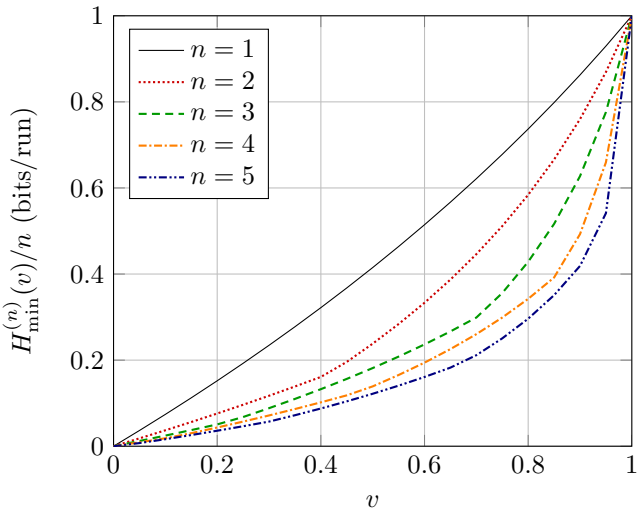


Figure 3.8: Min-entropy rates obtained when Eve is able to guess individual outcomes bits, and pairs, triples, quadruples and quintuples of outcome bits, in the ABNS scenario.

guessed from (3.2). This suggest, in analogy with other measures in quantum information, an asymptotic definition $\lim_{n\to\infty} \frac{H_{\min}^{(n)}(v)}{n}$ of the randomness of noisy PR-correlations in the ABNS, TONS, and WTONS scenarios.

3.5 Conclusion

In this Chapter, we have investigated the randomness of n noisy PR-boxes, which represent the paradigmatic example of nonlocal correlations and which are at the basis of many device-independent random number generation and quantum key distribution protocols.

In the Full-NS scenario, where the n noisy PR-correlations are obtained from n pairs of – possibly correlated – but non-communicating devices, the probability to guess correctly the n output bits of one party decreases exponentially with n, exactly as if the n noisy PR-boxes were n independent coins with a bias given by Eq. (3.2).

However, in the ABNS, TONS, and WTONS scenarios, where the n noisy

PR-correlations originate from a single pair of devices, either used sequentially n times, or which produce n outcome bits in one run, we have found that the randomness per PR-box can be significantly less than the individual randomness (3.2). In particular, we have found that, in the ABNS case, for noise values v below some threshold, the total randomness associated to $n \leq 5$ noisy PR-boxes is equal to the randomness of a single noisy PR-box. We conjecture that this holds for any n for some suitable noise threshold. In the TONS and WTONS case, for the same values of n, we have found that the guessing probability is linear in v for some region $[v_c^n; 1]$. We conjecture that this holds for any n, but that v_c^n tends to 1.

Besides their fundamental interest, it is worth considering our results from the perspective of the current status of the security of device-independent random number generation and quantum key distribution protocols. In the Full-NS scenario, their security has been proven [MRC+14]. In the case of the ABNS scenario, there exists a no-go result: starting from n noisy PR-boxes, it is not possible to extract, after privacy amplification, even a single bit that is arbitrarily close to uniform no matter how large n is [HRW13] (except if no noise is present, corresponding to $v = 1$). In the case of the TONS and WTONS scenarios, the situation is less clear. Though there exist severe limitations on the randomness one can extract from n noisy PR-boxes after privacy amplification [AFTS12], those results do not imply that DI RNG or QKD are necessarily impossible in these scenarios.

Interestingly, the Full-NS scenario, where security has been established, corresponds to the situation where the randomness of n noisy PR-boxes accumulates with n in an i.i.d. way, while in the ABNS, TONS, and WTONS scenarios, were security was proven to be impossible or is still an open question, the randomness per use of the PR-boxes decreases with n. Though the negative results that are presently known for the ABNS, TONS, and WTONS scenarios [AFTS12] are obtained by taking into account limitations on privacy amplification in a no-signalling context, it is possible that these impossibility results can be traced back to a lack of randomness even before privacy amplification.

To answer this question definitely, one would have to show that the smooth min-entropy is bounded by a sublinear (i.e. logarithmic or constant) function of n. The upper-bounds that we have obtained here are only concerned with the min-entropy, and thus do not imply any such impossibility result. Nevertheless, we believe that they pave the way to a new approach for studying the possibility of no-signalling privacy amplification, as no results were known concerning the min-entropy (smooth or non-smooth) in that context. Though our results do not exclude, in the ABNS, TONS and WTONS scenarios, a linear increase of

the min-entropy in the asymptotic limit $n \to \infty$, they imply an increase at a rate that is significantly lower than what one would naively deduce from the single-copy value (3.2).

Chapter 4

Regularising data for practical randomness generation

Non-local correlations that obey the no-signalling principle contain intrinsic randomness. In particular, for a specific Bell experiment, one can derive relations between the amount of randomness produced, as quantified by the min-entropy of the output data, and its associated violation of a Bell inequality. In practice, due to finite sampling, certifying randomness requires the development of statistical tools to lower-bound the min-entropy of the data as a function of the estimated Bell violation. The quality of such bounds relies on the choice of certificate, i.e., the Bell inequality whose violation is estimated. In this Chapter, we propose a method for choosing efficiently such a certificate. It requires sacrificing a part of the output data in order to estimate the underlying correlations. Regularising this estimate then allows one to find a Bell inequality that is well suited for certifying practical randomness from these specific correlations. We then study the effects of various parameters on the obtained min-entropy bound and explain how to tune them in a favourable way. Lastly, we carry out several numerical simulations of a Bell experiment to show the efficiency of our method: we nearly always obtain higher min-entropy rates than when we use a pre-established Bell inequality, namely the Clauser-Horne-Shimony-Holt inequality.

4.1 Introduction

In order to characterise the unpredictability of the outcomes of a given experiment, one usually models an adversary who has access to some information on the devices used in the experiment. If the devices in use behave classically, and if the adversary is given total information about them, no unpredictable bits can be obtained, as classical physics is deterministic. By contrast, if the devices are quantum, their outputs can be impossible to predict, even when the adversary has access to a perfect characterisation of the devices. In practice, a perfect control of quantum devices is rarely possible. This means that, in most cases, even the users do not have access to a perfect characterisation of the devices. Fortunately, the unpredictability of a sequence of bits can be certified even when the devices producing them cannot be completely characterised, thanks to the device-independent approach to quantum information protocols.

Quantifying the unpredictability of the bits obtained in a Bell experiment is not a trivial task, as it depends on a number of factors, including how powerful the adversary is assumed to be [VV12], how the devices are assumed to behave with time [AFTS12, RBaH$^+$16] or how the users process the accessible information [BSS14, NSBSP18]. In this Chapter, we adopt the most common approach to estimating the unpredictability of a Bell experiment: a user enters a bit in each of two shielded devices, which in return give output bits, according to some conditional probability distribution. These bits can be used to compute the observed violation of a Bell inequality. We compute the guessing probability given this violation, and we restrict our attention to the case where the adversary only has access to classical side information [FGS13, KZB17, NSBSP18] (for the case of an adversary with quantum side information, we refer the readers to [DFR16, KZF18, AFDF$^+$18]). That is a reasonable level of security since device-independent randomness generation involves only one user in one location. The only thing the adversary may exploit in this case is the imperfection of the device such as noise or deterioration with time. We refer the readers to [PM13] for a detailed explanation. We then quantify the randomness of the sequence of output bits by its min-entropy.

The upside of this approach is its simplicity, as it depends on only one parameter: the violation of a Bell inequality. However, in a real Bell experiment, this number cannot be exactly known, as the number of runs is finite. One can only compute an estimate of the average Bell violation. To overcome this obstacle, statistical tools were developed that allow one to upper-bound the predictability of the outputs with arbitrary confidence, based only on an estimate of the Bell

violation, rather than its theoretical value [PAM+10, FGS13, PM13, NSBSP18].

Another question naturally arises in this approach: which Bell inequality should one use to obtain good bounds? A Bell inequality violation contains only partial information about the input-output correlation. Choosing the inequality poorly can result in a serious underestimation of the unpredictability of a Bell experiment, and may not even certify any unpredictability, as every non-local correlations satisfy some Bell inequalities. Yet, if the input-output distribution is known, finding the Bell inequality that certifies as much randomness as possible turns out to be an SDP [NSPS14, BSS14]. Unfortunately, as mentioned above, the input-output distribution is not accessible in practice, due to finite statistics.

We thus propose a method to circumvent this problem. It consists in using part of the input-output statistics to estimate the corresponding underlying distribution. It is however very likely that a naive estimate based on the relative frequencies will not correspond to a distribution achievable with quantum physics. Consequently, the above-mentioned SDP is not directly applicable as it can only be solved for distribution that belongs to the quantum set, or to some specific relaxation of this set, such as the ones defined by the NPA hierarchy. We thus employ the methods developed in [LRZ+18] in order to obtain a distribution approximating the underlying distribution that lies inside one the NPA sets. This then enables us to solve the corresponding SDP and hence obtain a Bell inequality specifically suited for the estimated distribution, and hence better tailored for the underlying distribution.

4.2 Lower bound on the min-entropy

Alice and Bob perform n successive Bell tests. They respectively input $\mathbf{x} \in \{0,1\}^n$ and $\mathbf{y} \in \{0,1\}^n$ and obtain outputs $\mathbf{a} \in \{0,1\}^n$ and $\mathbf{b} \in \{0,1\}^n$, distributed according to the conditional distributions $P_{\mathbf{AB}|\mathbf{XY}}$. We assume that this behaviour obeys quantum mechanics. We quantify the randomness of the outputs produced in these Bells test via the min-entropy. We extend the definition given in Eq. (2.34), taking into account the possibility of conditioning on the observation of some event λ. The min-entropy of $(\mathbf{A}, \mathbf{B})$ given $(\mathbf{X}, \mathbf{Y})$ conditioned on some event λ, according to a distribution $P = P_{\mathbf{ABXY}}$, is:

$$H_{\min}(\mathbf{A}, \mathbf{B}|\mathbf{X}, \mathbf{Y}, \lambda)_P = -\log_2 \sum_{\mathbf{x},\mathbf{y}} P(\mathbf{x}, \mathbf{y}|\lambda) \max_{\mathbf{a},\mathbf{b}} P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}, \lambda). \tag{4.1}$$

The event λ is typically a function of the specific inputs that were chosen and the specific outputs that were obtained during the Bell experiment, such as a

statistical estimate. Note that, here, we don't take into account the adversary's side information E, because we assume that this side information is classical (see [FGS13, NSBSP18]). Moreover, we make the dependence of $H_{\min}$ on P explicit, as we will evaluate $H_{\min}$ on another distribution. We now introduce all the elements that allow us to lower-bound this quantity.

4.2.1 Statistical estimates

We first explain how Alice and Bob estimate the behaviour that underlies the Bell tests and the associated Bell violation. For a given Bell expression:

$$\mathcal{I}(P_{AB|XY}) = \sum_{a,b,x,y} c_{abxy} P_{AB|XY}(ab|xy), \tag{4.2}$$

$I_{\mathcal{L}}$ denotes its local bound, $\mathcal{I}(\mathcal{Q})$ denotes the interval of values achievable with quantum behaviours, $I_{\mathcal{Q}}^{+}$ denotes its maximal quantum value, and $I_{\mathcal{Q}}^{-}$, its minimal value:

$$\mathcal{I}(\mathcal{Q}) = \{\mathcal{I}(P) | P \in \mathcal{Q}\}, \qquad I_{\mathcal{Q}}^{+} = \max \mathcal{I}(\mathcal{Q}), \qquad I_{\mathcal{Q}}^{-} = \min \mathcal{I}(\mathcal{Q}). \tag{4.3}$$

For simplicity, we assume that the inputs $(\mathbf{x}, \mathbf{y})$ are chosen independently and identically at each round with probability $P(X_i = x, Y_i = y) = \pi_{xy}$.

Definition 9. *For a given realisation of* $(\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Y})$*, the observed frequencies* $\hat{P}_{AB|XY}$ *are defined as:*

$$\hat{P}_{AB|XY}(ab|xy) = \frac{N_{abxy}}{N_{xy}}, \tag{4.4}$$

where N_{abxy} *(resp.* N_{xy}*) is the number of occurrences of the quadruplet* (a, b, x, y) *(resp. the pair* (x, y)*) in the* n *length sequence* $(\mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y})$*.*

Definition 10. *For a given Bell expression* (4.2) *and a given realisation of* $(\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Y})$*, the observed average Bell violation* $\hat{I}$ *is:*

$$\hat{I} = \sum_{a,b,x,y} c_{abxy} \frac{N_{abxy}}{n \cdot \pi_{xy}}. \tag{4.5}$$

We point out that, even though $\hat{P}$ and $\hat{I}$ are both estimators, they do not involve the inputs in the same manner. To compute $\hat{P}$, one counts the occurrences of both the quadruplets (a, b, x, y) and the input pairs (x, y), whereas for $\hat{I}$, one

only counts the quadruplets (a, b, x, y) and uses directly the input distributions π_{xy}, instead of the frequencies of each input pair for a given realisation. Both can be computed from a realisation of Bell experiments, as π_{xy} is chosen by the user (see details hereafter). However, we decide to compute the observed frequencies in this way to ensure that $\hat{P}_{AB|X=x,Y=y}$ is normalised for each (x, y), and can thus be identified as a probability distribution. On the other hand, we decide to compute the observed Bell violation $\hat{I}$ directly using the input distribution, as this is crucial for the derivation of Theorem 5 (see [PM13, NSBSP18] for details). Note that, if the behaviours of the devices at each round are independent and identically distributed (i.i.d.) according to some distribution $P_{AB|XY}$, $\hat{I}$ converges towards $\mathcal{I}(P_{AB|XY})$ when n tends to infinity. However, we do not need to make such an assumption to define this quantity.

4.2.2 Randomness-bounding function

The key element that we use to lower bound the min-entropy (4.1) is a randomness-bounding function.

Definition 11. *Let χ be a subset of $\{0,1\}^2$. We say that $H_{\mathcal{I}}^{\chi} : \mathcal{I}(\mathcal{Q}) \to [0, 2]$ is a randomness-bounding function (RB function) for χ if it satisfies the two following requirements:*

R.1 $\forall P \in \mathcal{Q}$, $\min\limits_{\substack{(a,b)\in\{0,1\}^2 \\ (x,y)\in\chi}} (-\log_2 P(ab|xy)) \geq H_{\mathcal{I}}^{\chi}(\mathcal{I}(P))$,

R.2 $H_{\mathcal{I}}^{\chi}$ *is convex.*

These requirements are needed in order to bound the min-entropy produced by a sequence of Bell tests (see [NSBSP18] for a detailed explanation). Here, χ specifies a subset of all possible inputs for which the RB function is valid. It should contain the inputs for which the associated conditional distributions are the most random, i.e., the inputs that yield the largest $H_{\mathcal{I}}^{\chi}$. For instance, if one obtains a high $H_{\mathcal{I}}^{\chi}$ from one pair of input (x^*, y^*), and a small $H_{\mathcal{I}}^{\chi}$ for the others, one would have an interest in setting χ to (x^*, y^*) only. Indeed, the space over which the minimisation is carried out gets bigger when one includes more input pairs in χ, which results in a smaller RB function, which, in turn, will give a smaller lower bound on the min-entropy. However, this trade-off depends on the total number of Bell tests that are used for generating randomness, as is illustrated by the numerical simulations presented hereafter.

We now explain how to compute an RB function via the guessing probability problem. This general form was introduced and extensively explained in

[NSBSP18]. For a given Bell expression $\mathcal{I}$ and a specific value I^* of $\mathcal{I}$, finding the lower bound $H^\chi_\mathcal{I}(I^*)$ defined by requirements R.1 and R.2 amounts to solving a minimisation problem over all quantum behaviours P such that $\mathcal{I}(P) = I^*$. However, the optimisation problem obtained in this way is not easily solvable, due to the presence of the logarithm and to the complicated nature of the quantum set $\mathcal{Q}$ [NPA07, NPA08, GKW$^+$18].

This led the authors of [NSBSP18] to consider instead the following problem. For $(\alpha, \beta) \in \{0,1\}^2$ and $(\gamma, \delta) \in \chi$, let $\{\tilde{P}^{\alpha\beta\gamma\delta}\}$ be $4|\chi|$ variables, that represent unnormalised behaviours. The problem then reads:

$$\begin{aligned} G^\chi_\mathcal{I}(I^*) = \quad & \max_{\{\tilde{P}^{\alpha\beta\gamma\delta}\}} \quad \sum_{\substack{\alpha,\beta\in\{0,1\}^2 \\ \gamma,\delta\in\chi}} \tilde{P}^{\alpha\beta\gamma\delta}(\alpha\beta|\gamma\delta) \\ & \text{s.t.} \quad \sum_{\substack{\alpha,\beta\in\{0,1\}^2 \\ \gamma,\delta\in\chi}} \mathcal{I}(\tilde{P}^{\alpha\beta\gamma\delta}) = I^*, \\ & \quad \sum_{\substack{\alpha,\beta\in\{0,1\}^2 \\ \gamma,\delta\in\chi}} \mathrm{Tr}[\tilde{P}^{\alpha\beta\gamma\delta}] = 1, \\ & \quad \forall\ \alpha,\beta,\gamma,\delta,\ \tilde{P}^{\alpha\beta\gamma\delta} \in \tilde{\mathcal{Q}}_k, \end{aligned} \tag{4.6}$$

where $\mathrm{Tr}[\tilde{P}] = \sum_{ab} \tilde{P}(ab|xy)$ is the norm of $\tilde{P}$ (which is independent of (x, y) by no-signalling) and $\tilde{\mathcal{Q}}_k$ is the set of unnormalised behaviours that belong to the k^{th} level of the NPA hierarchy. This problem is then a an SDP. Moreover, if we let $H^\chi_\mathcal{I} = -\log_2 G^\chi_\mathcal{I}$, $H^\chi_\mathcal{I}$ satisfies both requirements R.1 and R.2, and is thus a RB function for χ (see [NSBSP18] for details). It is, however, not necessarily tight.

In the case where χ contains only one input pair (x^*, y^*), the guessing probability problem reduces to the conventional guessing probability $G(AB|x^*, y^*)$ presented in Eq. (2.27). In the same way as (2.27) could be extended to (2.28) to take into account the full underlying behaviour instead of only its value for a given Bell expression, one can define G^χ_{full} as:

$$\begin{aligned} G^\chi_{full}(P) = \quad & \max_{\{\tilde{P}^{\alpha\beta\gamma\delta}\}} \quad \sum_{\substack{\alpha,\beta\in\{0,1\}^2 \\ \gamma,\delta\in\chi}} \tilde{P}^{\alpha\beta\gamma\delta}(\alpha\beta|\gamma\delta) \\ & \text{s.t.} \quad \sum_{\substack{\alpha,\beta\in\{0,1\}^2 \\ \gamma,\delta\in\chi}} \tilde{P}^{\alpha\beta\gamma\delta} = P, \\ & \quad \forall\ \alpha,\beta,\gamma,\delta,\ \tilde{P}^{\alpha\beta\gamma\delta} \in \tilde{\mathcal{Q}}_k. \end{aligned} \tag{4.7}$$

As Problem (4.7) is more constrained than Problem (4.6), $G^{\chi}_{full}(P) \leq G^{\chi}_{\mathcal{I}}(\mathcal{I}(P))$. Moreover, as for (2.28), the dual problem of (4.7) returns a Bell expression $\mathcal{I}^*$ such that $G^{\chi}_{\mathcal{I}^*}(\mathcal{I}^*(P)) = G^{\chi}_{full}(P)$. When the Bell expression is well chosen, (4.6) and (4.7) are thus equivalent.

Let us stress however that these quantities can only be considered as theoretical measures of randomness for theoretical objects such as probability distributions and Bell expressions. In order to obtain practical bounds, one has to develop statistical tools.

4.2.3 Bounding the n round min-entropy

With the concepts defined above, we are now able to formulate a probabilistic statement on the min-entropy of the outputs obtained after a sequence of n Bell tests. Most of this section is a reformulation, adapted to our case, of the results first presented in [PAM+10], corrected in [PM13, FGS13], and extended in [NSBSP18]. Let us fix a behaviour $P_{\mathbf{AB}|\mathbf{XY}}$, an i.i.d. input distribution π_{xy}, and a Bell expression $\mathcal{I}$. The formal statement then reads:

Theorem 5. *Let $\{J_m | m \in [0, M]\}$ be a sequence of $M+1$ Bell violation thresholds, with $I_{\mathcal{L}} = J_0 < J_1 < ... < J_M = I^+_{\mathcal{Q}}$. Let λ_m be the event that the estimated Bell violation $\hat{I}$ falls between the thresholds J_m and J_{m+1}, and let $\mathbb{P}_{\tilde{P}}(\lambda_m)$ be the probability that this event occurs according to some distribution $\tilde{P}_{\mathbf{ABXY}}$. Let ϵ and ϵ' be two positive parameters. Then the true distribution $P_{\mathbf{ABXY}}$ is ϵ-close (with respect to the variational distance) to a distribution $\tilde{P}_{\mathbf{ABXY}}$ such that exactly one of these two statements holds:*

1. $\mathbb{P}_{\tilde{P}}(\lambda_m) \leq \epsilon'$,
2. $H_{min}(\mathbf{A}, \mathbf{B}|\mathbf{X}, \mathbf{Y}, \lambda_m)_{\tilde{P}_{\mathbf{ABXY}}} \geq n H^{\chi}_{\mathcal{I}}(J_m - \mu) - \gamma(\mathbf{x})\eta - \log_2 \frac{1}{\epsilon'}$,

where

$$\mu = \nu \sqrt{\frac{2}{n} \ln \frac{1}{\epsilon}}, \tag{4.8}$$

$$\nu = \max\{\max_{a,b,x,y} \frac{c_{abxy}}{\pi_{xy}} - I^-_{\mathcal{Q}}, I^+_{\mathcal{Q}} - \min_{a,b,x,y} \frac{c_{abxy}}{\pi_{xy}}\}, \tag{4.9}$$

$$\gamma(\mathbf{x}) = n - \sum_{i=1}^{n} \mathbb{1}_{\chi}(x_j), \tag{4.10}$$

$$\eta = \max\{H^{\chi}_{\mathcal{I}}(I^+_{\mathcal{Q}}), H^{\chi}_{\mathcal{I}}(I^-_{\mathcal{Q}})\}, \tag{4.11}$$

and $\mathbb{1}_\chi(x_j)$ *is the indicator function, which returns 1 if* $x_j \in \chi$ *and vanishes otherwise.*

The proof can be found in [PM13, NSBSP18]. Note that, unlike [NSBSP18], we take into account only one Bell expression in the statement of the theorem. This leads to numerous simplifications in its formulation, due in particular to the monotonicity of $H_{\mathcal{I}}^{\chi}$ over $[I_{\mathcal{L}}, I_{\mathcal{Q}}^{+}]$. In this sense, it is closer to the way it is stated in [PM13]. However, from [NSBSP18], we keep a few improvements on the parameters, and the possibility to select only a subset of inputs via χ. This enables improvement on the bound in some cases where the inputs have very different output probabilities: if the RB function is significantly better for a subset of inputs χ, this formulation allows to use the RB function for χ only, and corrects the bound via the penalty term $\gamma(\mathbf{x})\eta$. In that case, we have an interest in biasing the input distribution towards χ, in order to reduce the effect of the term $\gamma(\mathbf{x})\eta$ and thus produce as much randomness as possible. However, the trade-off between the quality of the RB function and the number of inputs from which randomness is generated depends on the total number of runs of a given protocol.

The bound given in the second statement of the theorem is the figure of merit that we aim at optimising in this work. Indeed, this expression depends on the choice of the Bell expression $\mathcal{I}$, and we now present a systematic approach to finding a well suited $\mathcal{I}$.

4.3 Results

We first present our new method for lower-bounding the min-entropy of the outputs of an uncharacterised Bell experiment. We then study, on a few behaviours, how the regularisation method, the size of sacrificed data, and the input distributions impact the quality of the min-entropy bound. We conclude by giving numerical results that illustrate the efficiency of our method.

4.3.1 Optimising the Bell expression via regularisation

As previously mentioned, solving the dual problem of (4.7) provides the Bell expression that is optimal for certifying the randomness of the given behaviour. When given an uncharacterised pair of devices, one could thus first generate some input-output data in order to estimate the corresponding underlying behaviour. This estimate $\hat{P}$ can then be used to obtain a Bell inequality that is presumably better for witnessing the randomness generated from these devices, by

computing the dual solution to the guessing probability problem. Unfortunately, as mentioned above, the guessing probability problem is only properly defined over the set of quantum behaviour $\mathcal{Q}$, or one of its NPA relaxation sets $\mathcal{Q}_k$, or over the set of no-signalling behaviours. On the other hand, there is no guarantee that the observed frequencies $\hat{P}$ belongs to any of these sets: $\hat{P}$ is on the contrary almost always signalling, even if the underlying behaviour is not, due to finite statistics. In this case, Problem (4.7) will be infeasible.

We now introduce our method to circumvent this problem, using the tools developed in [LRZ$^+$18]. The authors provided a set of tools to regularise the estimated behaviour $\hat{P}$ to one of the NPA sets $\mathcal{Q}_k$. It consists in minimising a norm-based metric or a statistical distance between $\hat{P}$ and $\mathcal{Q}_k$, the desired relaxation set, and taking the unique minimiser as the regularised behaviour $P^{\text{reg}}_{AB|XY}$. In this work, we employ two methods considered therein. The first one corresponds to minimising a statistical distance, namely the conditional Kullback-Leibler (KL) divergence [KL51, CJ06], and is defined in the following way:

$$P_{\text{ML}}(\hat{P}) = \underset{P\in\mathcal{Q}_k}{\text{argmin}}\, D_{\text{KL}}(\hat{P}||P), \tag{4.12}$$

where

$$D_{\text{KL}}(\hat{P}||P) = \sum_{a,b,x,y} \frac{N_{xy}}{n}\hat{P}(a,b|x,y)\log_2\Big(\frac{\hat{P}(a,b|x,y)}{P(a,b|x,y)}\Big).$$

and where ML stands for 'maximal likelihood'.

The second one corresponds to minimising the two-norm distance:

$$P_{\text{LS}}(\hat{P}) = \underset{P\in\mathcal{Q}_k}{\text{argmin}} \sqrt{\sum_{a,b,x,y}\Big(\hat{P}(a,b|x,y) - P(a,b|x,y)\Big)^2}, \tag{4.13}$$

where 'LS' stands for 'least-squares'. It is important to note that both these minimisations can be efficiently solved (see [LRZ$^+$18] for details), thus making this approach operationally relevant.

We can now define the following regularisation-based protocol for generating randomness from uncharacterised devices:

(i) Input a number N_{est} of (x,y) drawn from an i.i.d. uniform distribution (they can be public) and obtain the corresponding (a,b) in order to estimate the behaviour.

(ii) From this set of data, construct the observed frequencies $\hat{P}$ and compute $P^{\text{reg}}_{AB|XY}$, the regularisation of $\hat{P}$ (where $P^{\text{reg}}_{AB|XY}$ can be either $P_{\text{ML}}(\hat{P})$ or $P_{\text{LS}}(\hat{P})$).

(iii) Solve the corresponding optimisation problem $G^{\chi}_{full}(P^{\text{reg}}_{AB|XY})$ for different χ and select χ accordingly (see below for further details).

(iv) Extract the optimal Bell expression $\mathcal{I}$ from the dual.

(v) Input a number N_{raw} of (x, y), drawn according to a distribution P^{χ}_{XY} (they can be public), obtain the corresponding (a, b), and compute the observed Bell violation $\hat{I}$.

(vi) Apply Theorem 5 to lower-bound the min-entropy of the raw set of data $(a_i, b_i, x_i, y_i)_{i\in\{1,N_{\text{raw}}\}}$.

We now make a few observations on this protocol, which is summarised in Figure 4.1. The subset χ is chosen at step (iii), thanks to $P^{\text{reg}}_{AB|XY}$. Indeed, $P^{\text{reg}}_{AB|XY}$ reveals some information about the underlying behaviour. One might thus intuitively do the following: compute the values of $G^{(x,y)}_{full}(P^{\text{reg}}_{AB|XY})$ for all the inputs, and decide accordingly; if the value is roughly the same for all (x, y), one would choose $\chi = \{0, 1\}^2$; if one input pair (x^*, y^*) yields a lower guessing probability, one would choose $\chi = (x^*, y^*)$. However, if N_{raw} is not big enough, $\chi = \{0, 1\}^2$ is likely to result in a better min-entropy bound in any case, as our results show.

The optimised Bell expression $\mathcal{I}$ obtained in step (iv) may not be unique and the different possible representations of $\mathcal{I}$ are only artefacts of numerical computations. However, the choice of a representative for $\mathcal{I}$ matters, since two physically equivalent representations can lead to different statistical estimates [RRMG17], and thus to distinct lower bounds on the min-entropy. In order to avoid such effects, we use the unique representation introduced in [RRMG17], by setting the signalling part to zero (see [RRMG17] for details).

In step (v), we assume that the specific distributions P^{χ}_{XY} can be generated using some freely available resource. If this is the case, one might consider that the task of randomness generation is already achievable, and we might then call our primitive 'randomness expansion', rather than 'randomness generation'. However, the input randomness can be public: it needs to be random to anyone beforehand, but it can be accessed by anyone after it is produced. Conversely, the output randomness is private: its value resides in the fact that it is only

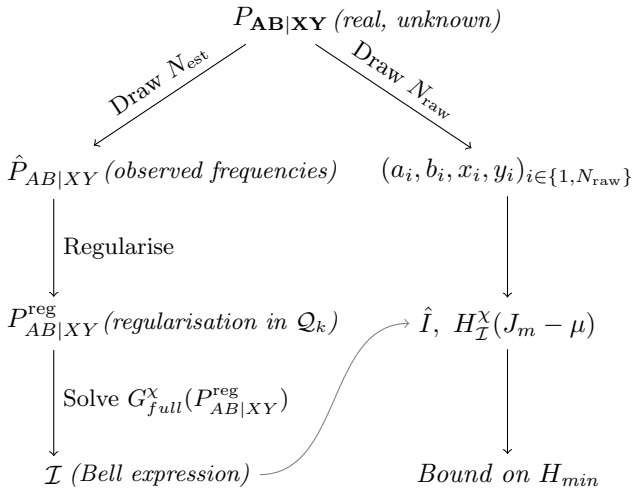


Figure 4.1: Schematic representation of our protocol: the user draws N_{est} bits from the unknown underlying behaviour, collects the frequencies and regularises them to obtain an estimate that lies in one of the NPA sets. The dual of the corresponding guessing probability problem provides a Bell inequality that is then used to quantify the min-entropy of the sequence of N_{raw} bits.

accessible to the user. We can thus refer to this process as 'private random bits generation'.

In step (vi), we only bound the min-entropy of the data generated in step (v). Indeed, it is essential that the set of data used for the estimation be different from the one for which the bound on the min-entropy is derived: the statistical analysis of the data cannot depend on the data itself. This implies that, contrarily to [NSBSP18], our method requires that part of the data is used only for parameter estimation, and then thrown away.

Finally, note that even though the regularisation method described in [LRZ$^+$18] is meaningful only when the underlying distribution $P_{\mathbf{AB|XY}}$ is i.i.d., the derivation of the bound on the min-entropy does not rely on this assumption. For this reason, the probabilistic statement that we obtain via our method will still be valid, even if $P_{\mathbf{AB|XY}}$ is not i.i.d.. In this case, the Bell expression that we obtain might be inadequate, which might result in a trivial lower bound on the min-entropy (that equals to zero), but it will not result in an overestimation

of the min-entropy of the raw data. In this sense, the optimisation method might become irrelevant, but the security analysis will not be compromised.

4.3.2 Tuning the parameters

In order to adjust the parameters of our protocol, we simulate some pairs of devices, by generating for each one a random state ρ and some random measurements $\{M^{\mathrm{A}}_{a|x}\}_a$ and $\{M^{\mathrm{B}}_{b|y}\}_b$. The random states are picked at random in the space of two qubit pure states via their Schmidt decomposition, and the random measurements are generated via their associated projectors, picked at random on the Bloch sphere.

We then compute the associated behaviour:

$$P_{AB|XY}(ab|xy) = \mathrm{Tr}[\rho\, M^A_{a|x} \otimes M^B_{b|y}]. \tag{4.14}$$

To ensure that the obtained behaviours are nonlocal, we compute their associated CHSH values $\mathcal{I}_{\mathrm{CHSH}}(P_{AB|XY})$:

$$\mathcal{I}_{\mathrm{CHSH}}(P_{AB|XY}) = \sum_{x,y,a,b} (-1)^{xy+a+b} P_{AB|XY}(ab|xy), \tag{4.15}$$

and discard those for which $\mathcal{I}_{\mathrm{CHSH}}(P_{AB|XY}) \leq 2$. We then construct the corresponding N_{tot}-round behaviour using $P_{AB|XY}$ in an i.i.d. way, i.e.,

$$P_{\mathbf{AB}|\mathbf{XY}}(\mathbf{ab}|\mathbf{xy}) = \prod_{i=1}^{N_{\mathrm{tot}}} P_{AB|XY}(a_i b_i|x_i y_i). \tag{4.16}$$

We set $N_{\mathrm{tot}} = N_{\mathrm{est}} + N_{\mathrm{raw}} = 10^8$, in accordance with the state-of-the-art experimental demonstration of device-independent randomness generation [BKG+18]. We then conduct a detailed study of four of these random behaviours, to heuristically fix three crucial parameters of our protocol:

- the regularisation method,
- the number of rounds used for the estimation N_{est},
- the inputs subset used to generate randomness χ.

Based on the data we obtained, presented in Appendix B.1, we decided to set:

- $P^{\text{reg}}_{AB|XY} = P_{ML}$,
- $N_{\text{est}} = 10^6$,
- $\chi = \{0,1\}^2$

The graphs that corroborate these decisions can be found in Appendix B.1. Before we give the results of several simulations that illustrate the efficiency of our protocol, note that, when one sets $N_{\text{tot}} = 10^8$, generating randomness from only one input pair (i.e., setting $\chi = (x^*, y^*)$) does not usually result in higher min-entropy bounds than when one sets $\chi = \{0,1\}^2$. The same effect can be observed in the simulations carried out by the authors in [NSBSP18]. It is not surprising: in order to obtain a good min-entropy rate when certifying randomness from only one input pair, one should bias the input distribution towards that pair as much as possible. However, in order to obtain a reliable estimate of the Bell violation, one should evaluate it with many occurrences of each possible input. These two assertions are in an apparent contradiction, and they can both hold simultaneously only if N_{tot} is high enough. It seems that, for most behaviours, $N_{\text{tot}} = 10^8$ is not sufficient. We however checked that, when N_{tot} is sufficiently big, our method provides better min-entropy bounds for $\chi = (x^*, y^*)$ than for $\chi = \{0,1\}^2$. The corresponding graph can be found in Appendix B.2.

4.3.3 Numerical results

Our figure of merit is the comparison between the min-entropy bound obtained from our protocol, denoted $H_{\min}$ in the following, and the one obtained from a direct evaluation of the CHSH inequality, $H^{\text{CHSH}}_{\min}$. We generate 50 behaviours at random (in the same way as described above) and run 500 simulations for each of them. To compute the lower bound on $H_{\min}$, one should set $n = N_{\text{raw}}$ in Theorem 5, whereas for $H^{\text{CHSH}}_{\min}$, $n = N_{\text{tot}} > N_{\text{raw}}$, as no estimation is required.[1]

The parameters of the bound of Theorem 5 are set as follow: we fix $\epsilon = \epsilon' = 10^{-6}$, we divide the interval $[I_{\mathcal{L}}, I^+_{\mathcal{Q}}]$ in $M + 1 = 1000$ segments of the same length, and we use the NPA local level 2 [MBL$^+$13] for the regularisation and the guessing probability problems. We then compute the corresponding

[1] It might seem necessary to also first sacrifice a part of the data to determine which among the 8 representatives of the CHSH inequality is violated. This is however unnecessary as any given behaviour can violate at most one representative of the CHSH inequality (see page 2 of the Supplementary Material to Ref [LHBR10]), which can be determined by evaluating the min-entropy bound of all different representatives of the CHSH inequality.

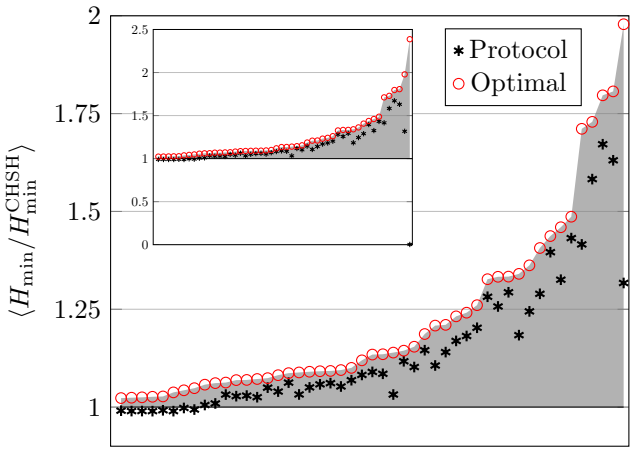


Figure 4.2: Black asterisk: ratio between the rate obtained via our protocol and via the direct use of the CHSH inequality. Red circle: ratio between the maximal achievable min-entropy and the rate obtained via the direct use of the CHSH inequality. The inset contains all 50 simulations, including the single instance from which no randomness is certified (see explanation in the main text). This exceptional point is removed from the main plot so that the remaining (successful) cases can be examined more closely.

min-entropy rate by dividing these values by N_{tot} in both cases. We also computed $-\log_2(G^{\chi}_{full}(P_{AB|XY}))$, which corresponds to the maximal achievable min-entropy rate. To show that it is worth sacrificing part of the data for estimation, we then compared these three quantities. The results are presented in Fig. 4.2.

In this figure, we plot the ratios between the min-entropy rates for $H_{\min}$ and $H_{\min}^{\mathrm{CHSH}}$ for every simulated pairs of devices, as well as the ratios between the maximal achievable rate $-\log_2(G^{\chi}_{full}(P_{AB|XY}))$ and $H_{\min}^{\mathrm{CHSH}}$. For clarity, we sorted them in ascending order of the latter. We highlighted in grey the areas between the line $y = 1$, where the amount of randomness given by our protocol is the same as using CHSH inequality, and the curves connecting the optimal rates. Our protocol is good whenever a point falls in this area. Indeed, it means that, despite the N_{est} bits that were thrown away, we obtain a higher bound on the min-entropy than if we had directly used the CHSH inequality on all the bits.

We observe that our method performs well in 98% of the simulations, in the following sense: when the optimal rate is nearly achieved with the CHSH inequality (i.e., the CHSH inequality gives a bound that is above 95% of the optimal rate), so does our method; when the CHSH inequality does not achieve the optimal rate, our method performs significantly better (with rates up to 1.6 times more) in all but one case.

We now explain what happened with that last point of our simulations, from which no randomness can be certified via our protocol. The corresponding underlying behaviour has a low CHSH value, and the optimal Bell inequality is such that the gap between the local bound and the quantum bound is very small. This seems to indicate that this behaviour is of the kind presented in [AMP12], i.e., it is almost local, but also close to the border of the quantum set. The authors of [AMP12] proved that, in theory, a lot of randomness could be certified from such behaviours, as can be observed by the corresponding red circle in Fig. 4.2. However, those behaviours are not good from a practical point of view: the small gap between the local and the quantum bounds of their associated optimal Bell inequality requires that the confidence interval on the estimated Bell violation $J_m - \mu$ be very small. If not, no Bell violation can be observed, and thus no randomness can be certified. This is the case for that point of our simulations.

4.4 Conclusion

We presented a simple method to optimise the lower bound derived in [NSBSP18] on the min-entropy produced by a sequence of Bell tests. It consists in estimating the underlying behaviour of the black boxes, via the regularisation method given in [LRZ$^+$18]. We then tuned the parameters of this protocol via a heuristic method. We concluded that, when one regularises some data for randomness generation, one should always use the maximal likelihood method (the authors observed the same effect for another figure of merit, the negativity, in [LRZ$^+$18]), one can sacrifice up to 1% of the data for estimation, and that, for the device-independent randomness generation experiments that can be performed at the moment (i.e., with $N_{\text{tot}} = 10^8$), one should generally use the worst case RB function (i.e., the one that bounds the randomness for all inputs). We then carried out numerical simulations that illustrate the efficiency of this method.

We now describe two possible future lines of investigation. The first one would be to take into account more factors in the optimisation of the lower bounds on the min-entropy. For instance, one could generate randomness from

two or three subsets of inputs pairs, instead of considering only one or all of them as we did here. One could also tune P^{χ}_{XY} in a more precise way, as a function of the total number of rounds N_{tot} and of the differences between the guessing probabilities for each input pair. Finally, the RB function is a key element in the derivation of the bound. We used here the one introduced in [NSBSP18]. However, there are other ways to compute a function that satisfies both requirements R.1 and R.2 needed for an RB function, such as the one introduced in [BSS14]. Being able to compute the RB function that is tight would entail an improvement on the min-entropy bound.

The second one is related to the power given to the adversary. Our results hold in a trusted provider scenario, where our protocol allows for correcting noise and deterioration in the apparatuses, and in an adversarial scenario where the adversary holds only classical-side information. Adapting it to the case of an adversary with quantum side information would provide a min-entropy bound valid in the most general scenario. This could be achieved via the entropy accumulation theorem [DFR16]. Based on that result, a bound was derived on the n-round smooth min-entropy against an adversary with quantum side information [AFDF$^+$18]. However, this bound is based on the CHSH inequality (or, more accurately, on the CHSH game). Deriving such a bound for other inequalities might be a hard task. We took a different approach here, that consists in optimising the amount of randomness that is generated by tailoring the Bell inequality to a specific case. This, in turn, led us to consider only classical side information. If one could adapt the results of [DFR16, AFDF$^+$18] to any Bell inequality, one would be able to guarantee the security of our protocol in the most general scenario.

Chapter 5

Randomness versus non locality in the Mermin-Bell experiment with three parties

The detection of nonlocal correlations in a Bell experiment implies almost by definition some intrinsic randomness in the measurement outcomes. For given correlations, or for a given Bell violation, the amount of randomness predicted by quantum physics, quantified by the guessing probability, can generally be bounded numerically. However, currently only a few exact analytic solutions are known for violations of the bipartite Clauser-Horne-Shimony-Holt Bell inequality. In this Chapter, we study the randomness in a Bell experiment where three parties test the tripartite Mermin-Bell inequality. We give tight upper bounds on the guessing probabilities associated with one and two of the parties' measurement outcomes as a function of the Mermin inequality violation. Finally, we discuss the possibility of device-independent secret sharing based on the Mermin inequality and argue that the idea seems unlikely to work. The results of this Chapter are based on [WBA18a].

5.1 Introduction

The detection of nonlocal correlations in a Bell experiment implies some randomness in the measurement outcomes, regardless of the exact physical mechanism by which the correlations are produced, provided that communication between the sites is prohibited. The simplest measure of randomness and typically the easiest to bound is the guessing probability. Aside from its direct operational meaning, the guessing probability is a useful quantity in the analysis of device-independent cryptography protocols: security proofs of device-independent protocols frequently depend on a lower bound on the min-entropy (a function of the guessing probability) or the conditional von Neumann entropy (which the min-entropy is a lower bound for) [RGK05, Ren05, MPA11, PM13, AFDF$^+$18]. In the practically most relevant case where the measurements are made on a quantum system, a numeric method for deriving an upper bound on the guessing probability exists, based on the NPA hierarchy of relaxations of the optimisation problem to SDP [NPA07, NSPS14, BSS14], for which reliable optimisation algorithms exist.

Since the determination of guessing-probability bounds by numerical means is essentially a solved problem, our interest here is in cases where it is possible to establish a tight analytic bound. Currently, only a few tight bounds on the guessing probability are known for the Clauser-Horne-Shimony-Holt (CHSH) [CHSH69] inequality. As explained in Chapter 2.3.1, the adversary Eve's probability of guessing one of one party's (say, Alice's) measurement outcomes is equal to [PAM$^+$10]:

$$G(A|x=0) \leq \frac{1}{2}(1+\sqrt{2-S^2/4}) \tag{5.1}$$

for a given CHSH expectation value S. More recently, Kaniewski and Wehner [KW16] have derived the tight upper bound

$$G(A|B) \leq \frac{1}{2}+\frac{1}{4}\Big(S/2+\sqrt{2-S^2/4}\Big) \tag{5.2}$$

on an average probability $G(A|B) = \big(P(A=B|X=0,Y=2)+P(A=B|X=1,Y=2)\big)/2$ that the second party Bob is able to guess Alice's measurement outcome without knowing which measurement Alice performed, assuming they are chosen equiprobably.

Beyond the CHSH scenario, guessing-probability bounds have been determined for violations of bipartite and multipartite chained Bell inequalities [BKP06, AGCA12]; however these are derived assuming only the no-signalling

constraints and they are not generally tight assuming the scenario is restricted to correlations and attacks allowed by quantum physics.

Here, we study the amount of randomness that can be certified in a Bell experiment with three parties showing a violation of Mermin's tripartite Bell inequality [Mer90]. We report tight bounds for the following two cases:

- The guessing probability $G(A|x=0)$ associated with the measurement outcome at one site, in terms of two independent Mermin expectation values.
- The guessing probability $G(AB|x=0,y=0)$ associated with measurement outcomes at two sites, for a given violation of one Mermin inequality.

5.2 Scenario and results

Our results apply to the following adversarial Bell scenario: three cooperating parties, Alice, Bob, and Charlie, and an eavesdropper, Eve, share a quantum state ρ_{ABCE} on some Hilbert space $\mathcal{H}_{\mathrm{A}}\otimes\mathcal{H}_{\mathrm{B}}\otimes\mathcal{H}_{\mathrm{C}}\otimes\mathcal{H}_{\mathrm{E}}$. Alice, Bob, and Charlie may each perform one of two measurements indexed $x, y, z \in \{0,1\}$ on their part of the state, which yield respective outcomes $a, b, c \in \{+,-\}$. Eve performs a measurement yielding an outcome e, intended to be correlated with one or more of Alice's, Bob's, and Charlie's outcomes. Generally, we will assume, without loss of generality, that Eve's measurement has the same number of outcomes as the number of possible different results that the cooperating parties may obtain that she wishes to distinguish between. The joint correlations are summarised by a table of conditional probabilities

$$P(abce|xyz) = \mathrm{Tr}\left[\left(\Pi^{\mathrm{A}}_{a|x}\otimes\Pi^{\mathrm{B}}_{b|y}\otimes\Pi^{\mathrm{C}}_{c|z}\otimes\Pi^{\mathrm{E}}_{e}\right)\rho_{\mathrm{ABCE}}\right], \tag{5.3}$$

where $\Pi^{\mathrm{A}}_{a|x}$ is the measurement operator associated with the outcome a when Alice performs the measurement x, and similarly for Bob's, Charlie's, and Eve's measurement operators $\Pi^{\mathrm{B}}_{b|y}$, $\Pi^{\mathrm{C}}_{c|z}$, and Π^{E}_{e}. The measurements can be assumed to be projective, since we do not assume any limit on the dimension of the underlying Hilbert space. The state and measurements are all treated as unknown except possibly to Eve.

Eve's goal in this setting is to be able to guess one or more of Alice's, Bob's and/or Charlie's measurement outcomes. The simplest measure of her ability to do so, the guessing probability, is simply the probability that Eve's guess is correct. In the simplest case where Eve aims to guess (say) Alice's $x=0$

measurement outcome, the ("local") guessing probability is the probability that Eve's measurement outcome is the same as Alice's:

$$G(A|x=0) = \sum_a P_{AE|X}(A=a, E=a|X=0)\,, \tag{5.4}$$

where $P_{AE|X}(A=a, E=e|x) = \sum_{bc} P(abce|xyz)$ (see Eq. (2.25)). Other guessing probabilities are straightforward variations of this. For instance, the guessing probability associated with Alice's and Bob's joint outcomes for measurements $x=y=0$ is:

$$G(AB|x=0, y=0) = \sum_{a,b} P_{ABE|XY}\big(ab(a,b)|X=0, Y=0\big)\,, \tag{5.5}$$

where we label Eve's (four) possible measurement outcomes $(++)$, $(+-)$, $(-+)$, and $(--)$. Alice, Bob, and Charlie wish to certify that Eve's ability to guess outcomes is limited (in mathematical terms, that guessing probabilities like (5.4) and (5.5) must be less than one) using only the information available to them, encapsulated by the marginal distribution $P_{ABC|XYZ}(abc|xyz) = \sum_e P(abce|xyz)$. A necessary but not necessarily sufficient condition for this is that this marginal distributions does not admit a local hidden variable model, i.e., it does not admit a factorisation of the form

$$P_{ABC|XYZ}(abc|xyz) = \sum_\lambda p_\lambda P_{A|X\Lambda}(a|x,\lambda) P_{B|Y\Lambda}(b|y,\lambda) P_{C|Z\Lambda}(c|z,\lambda)\,, \tag{5.6}$$

which is detected if the marginal distributions $P_{ABC|XYZ}$ violate a Bell inequality.

Here, we study the amount of randomness that can be certified in this tripartite scenario if a violation of the Mermin-Bell inequality is observed. The Mermin inequality [Mer90] $M \leq 2$ holds for local-hidden-variable models, where the Mermin correlator is

$$M = \langle A_0 B_0 C_0\rangle - \langle A_0 B_1 C_1\rangle - \langle A_1 B_0 C_1\rangle - \langle A_1 B_1 C_0\rangle\,, \tag{5.7}$$

and in turn $\langle O\rangle$ denotes the expectation value of the observable quantity O. In the quantum case, $\langle O\rangle = \mathrm{Tr}[O\rho_{\mathrm{ABC}}]$ is given by the expectation value in the underlying marginal state ρ_{ABC} and the dichotomic operators $-\mathbb{1} \leq A_x, B_y, C_z \leq \mathbb{1}$ are related to the measurement operators by

$$A_x = \Pi^{\mathrm{A}}_{+|x} - \Pi^{\mathrm{A}}_{-|x}, \quad B_y = \Pi^{\mathrm{B}}_{+|y} - \Pi^{\mathrm{B}}_{-|y}, \quad C_z = \Pi^{\mathrm{C}}_{+|z} - \Pi^{\mathrm{C}}_{-|z}. \tag{5.8}$$

The Mermin inequality is best known for its association with the Greenberger-Horne-Zeilinger (GHZ) paradox [GHSZ90]. The maximal quantum (and algebraic) violation $M = 4$ is attained by measuring $A_0 = B_0 = C_0 = \sigma_x$ and $A_1 = B_1 = C_1 = \sigma_y$ on the GHZ state $|\Psi\rangle = (|111\rangle + |222\rangle)/\sqrt{2}$. Violations greater than $2\sqrt{2}$ require entanglement between all three sites [BGLP11].

The Mermin expression M can be obtained as the real part of the quantity

$$\big\langle (A_0 + iA_1)(B_0 + iB_1)(C_0 + iC_1) \big\rangle \,. \tag{5.9}$$

The imaginary part is also a Mermin expression,

$$M' = \langle A_0 B_0 C_1 \rangle + \langle A_0 B_1 C_0 \rangle + \langle A_1 B_1 C_0 \rangle - \langle A_1 B_1 C_1 \rangle \,, \tag{5.10}$$

equivalent to (5.7) up to relabelling some of the inputs and outputs. The sum $M_+ = M + M'$ is the correlator appearing in Svetlichny's inequality [Sve87], which was constructed to always require nonlocality (and thus entanglement) between all three parties in order to violate.

Some randomness bounds, quantified by guessing probabilities involving one, two, and three parties, are illustrated in figures 5.1 and 5.2 in terms of the Mermin and Svetlichny expectation values. Of these, we were able to find the analytic form of the curve for the local guessing probability $G(A|0)$ in both cases and the curve for $G(AB|00)$ in terms of the Mermin expectation value.

For given values of the Mermin or Svetlichny correlators, the corresponding upper bounds on the local guessing probability have the same functional form,

$$G(A|0) \leq f(M) \qquad \text{and} \qquad G(A|0) \leq f\big(M_+/\sqrt{2}\big) \,, \tag{5.11}$$

for the function

$$f(x) = \begin{cases} \frac{1}{2} + \frac{1}{2}\sqrt{x(1 - x/4)} & \text{if } x \geq 2 + \sqrt{2} \\ 1 + \frac{1}{\sqrt{2}} - x/4 & \text{if } x \leq 2 + \sqrt{2} \end{cases} \tag{5.12}$$

in the range $2\sqrt{2} \leq x \leq 4$. Both are implied by the tight bound

$$G(A|0) \leq f\big(\sqrt{M^2 + M'^2}\big) \,, \tag{5.13}$$

in which the two Mermin expectation values M and M' appear as independent parameters. Note that since f is a decreasing function in its argument, (5.13) is equivalent to stating that

$$G(A|0) \leq f\big(\cos(\varphi) M + \sin(\varphi) M'\big) \tag{5.14}$$

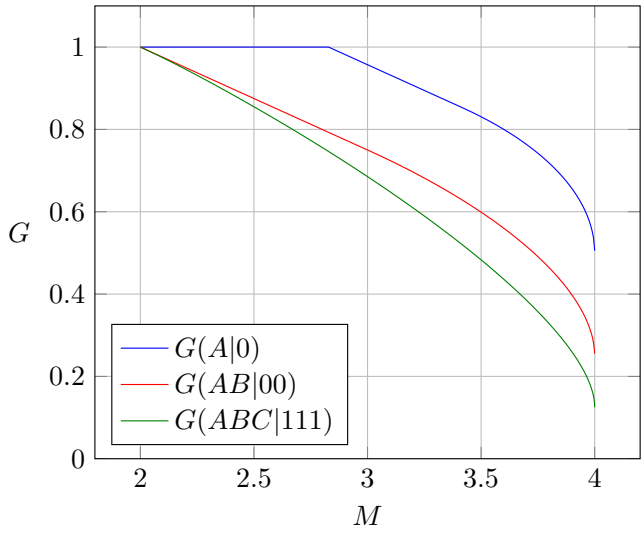


Figure 5.1: Upper bounds on the guessing probabilities $G(A|0)$, $G(AB|00)$, and $G(ABC|111)$ for expectation values $2 \leq M \leq 4$ of the Mermin expression. The upper bound for $G(ABC|111)$ was determined numerically at the level $1 + \mathrm{A}^2 + \mathrm{AB} + \mathrm{AC} + \mathrm{BC}$ of the NPA hierarchy.

holds for all φ. The result (5.13) certifies some intrinsic randomness for values of M and M' satisfying

$$2\sqrt{2} < \sqrt{M^2 + M'^2} \leq 4\,. \tag{5.15}$$

For M alone and the Svetlichny combination $M_+ = M + M'$, randomness for one measurement outcome is certified for $M > 2\sqrt{2}$ and $M_+ > 4$. This is what one would expect, since these are precisely the ranges that require entanglement between all three parties to attain. At the boundary $\sqrt{M^2 + M'^2} = 4$, (5.13) reduces to $G(A|0) \leq 1/2$, certifying that the measurement outcome must be uniformly random.

In the case that the eavesdropper aims to jointly guess two parties' measurement outcomes, the guessing probability respects the tight bound

$$G(AB|00) \leq \begin{cases} \frac{3}{4} - \frac{M}{8} + \sqrt{3}\sqrt{\frac{M}{8}\left(\frac{1}{2} - \frac{M}{8}\right)} & \text{if } M \geq 3 \\ \frac{3}{2} - \frac{M}{4} & \text{if } M \leq 3 \end{cases} \tag{5.16}$$

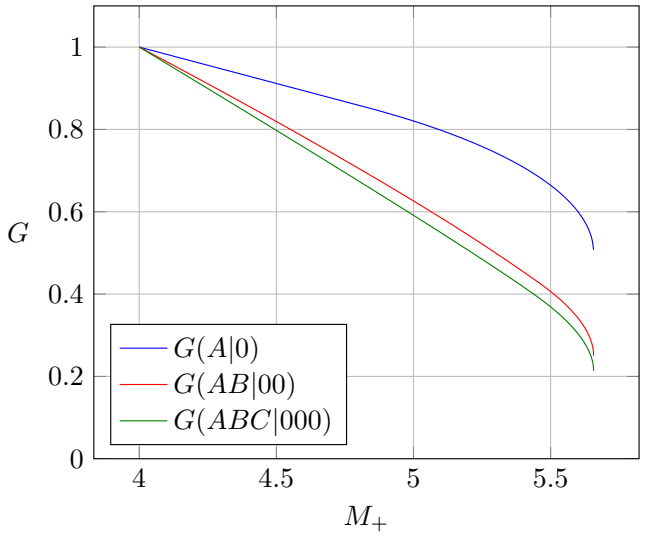


Figure 5.2: Guessing probabilities for expectation values $4 \leq M_+ \leq 4\sqrt{2}$ of the Svetlichny expression $M_+ = M + M'$. The upper bounds for $G(AB|00)$ and $G(ABC|000)$ were obtained numerically at levels $1 + \mathrm{AB} + \mathrm{AC} + \mathrm{BC}$ and $1 + \mathrm{A}^2 + \mathrm{AB} + \mathrm{AC} + \mathrm{BC}$ of the NPA hierarchy.

in the range $2 \leq M \leq 4$. In this case, we detect some randomness as soon as the local bound $M \leq 2$ is violated. The maximum possible violation $M = 4$ implies $G(AB|00) \leq 1/4$, corresponding to the maximum possible randomness.

Beyond this we did not find any new tight bounds for violations of the Mermin inequality. The upper bound for the global guessing probability $G(ABC|000)$ in terms of M is exactly the same as (5.16), while the upper bound for $G(ABC|111)$ (which should attain $1/8$ if the Mermin inequality is maximally violated [SG01, WW01]) appears to be the solution to the maximisation problem

$$\begin{aligned} \max \quad & \frac{1}{8}\Big(1 + 24\cos\big(\tfrac{3}{2}\theta_2\big)\alpha\beta + \ 2\cos(3\theta_2)\alpha^2 + 30\beta^2\Big) \\ \text{s.t.} \quad & M = \big(2\cos(3\theta_1) - 6\cos(\theta_1 + 2\theta_2)\big)\alpha^2 - 12\cos(\theta_1 - \theta_2)\beta^2 \\ & 2\alpha^2 + 6\beta^2 = 1 \end{aligned} \tag{5.17}$$

over $\alpha, \beta, \theta_1, \theta_2 \in \mathbb{R}$, which we were unable to significantly simplify further (let alone prove). Eq. (5.16) also does not generalise in terms of M and M' in the way that the local-guessing-probability bound does. The upper bound for $G(AB|00)$ in terms of the Svetlichny combination (illustrated in figure 5.2) for instance has a different form than (5.16). This is expected since the local-guessing-probability bound is already less than 1 for any violation of the local bound, and we were not much more successful in attempting to identify it analytically than we were for $G(ABC|111)$ in terms of M.

For simplicity we have stated the results (5.13) and (5.16) for the guessing probabilities $G(A|0)$ and $G(AB|00)$; however symmetries of the Mermin correlator(s) imply that the bounds are the same regardless of what measurements are considered. For the global guessing probabilities there are two inequivalent cases, $G(ABC|000)$ and $G(ABC|111)$, in terms of M.

In figures 5.1 and 5.2 we have also included upper bounds on guessing probabilities for which we do not have an exact analytic expression. We derived these numerically by solving the semidefinite programming relaxations at the levels of the NPA hierarchy indicated in the figure captions. We used the arbitrary-precision solver SDPA-GMP [SDP11, Nak10] for this purpose. We have made the code we used to generate the relaxations available online [Woo18].

5.3 Tangent Bell expressions

We have asserted that the local and two-party guessing probabilities respect the upper bounds (5.13) and (5.16) and that the bounds are tight. We prove these assertions in this section.

5.3.1 General idea illustrated with CHSH

Proving the main results (5.13) and (5.16) is equivalent to proving families of linear inequalities corresponding to tangents of the curves. We illustrate the approach using CHSH as an example, for which this has already been done [MPA11, AMP12]. It was shown in [AMP12] that the quantum expectation value of a modified CHSH expression respects the tight upper bound

$$\beta\langle A_0\rangle + S \leq 2\sqrt{2}\sqrt{1+\beta^2/4} \tag{5.18}$$

in the parameter range $0 \leq \beta \leq 2$. Eq. (5.18) can be rewritten as an upper bound

$$\langle A_0\rangle \leq \frac{1}{\beta}\Big(2\sqrt{2}\sqrt{1+\beta^2/4} - S\Big) \tag{5.19}$$

for $\langle A_0 \rangle$. Assuming that $S \geq 2$, minimising the right-hand side over β produces the tightest possible bound

$$\langle A_0 \rangle \leq \sqrt{2 - S^2/4}\,. \tag{5.20}$$

This bound has two key characteristics. First, since the CHSH expression remains unchanged under (for example) the replacements $A_x \mapsto -A_x$ and $B_y \mapsto -B_y$, the same upper bound holds for $-\langle A_1 \rangle$ as well as $\langle A_1 \rangle$. Second, the right-hand side is by construction concave in S. Using these properties and that $P_{A|X}(+|x) = (1 + \langle A_x \rangle)/2$ and $P_{A|X}(-|x) = (1 - \langle A_x \rangle)/2$, the result is quickly obtained:

$$\begin{aligned} G(A|0) &= \sum_a P_{AE|X}(aa|0) \\ &= \sum_a P_E(a) P_{A|XE}(a|0,a) \\ &\leq \sum_a P_E(a) \frac{1}{2}\Big(1 + \sqrt{2 - {S_{|a}}^2/4}\Big) \\ &\leq \frac{1}{2} + \frac{1}{2}\sqrt{2 - S^2/4}\,, \end{aligned} \tag{5.21}$$

where $S_{|a}$ in the third line is the CHSH expectation value conditioned on Eve obtaining the outcome $e = a$.

In passing, we mention that the bound (5.2) for $G(A|B) = \frac{1}{2} + \frac{1}{4}\langle (A_0 + A_1)B_2 \rangle$ can similarly be derived from the inequality

$$\alpha \langle (A_0 + A_1)B_2 \rangle + S \leq 2\sqrt{1 + (1+\alpha)^2} \tag{5.22}$$

for $\alpha \geq 0$. The inequality (5.22) itself is implied by the tight quantum bound derived for the I_α^β expression in [AMP12], since there is clearly no advantage for the operator B_2 to be different from B_0 in order to maximise the left-hand side.

The same general approach works for the main results of section 5.2. The Mermin expectation values M and M' are both symmetric under the transformations $A_x, C_z \mapsto -A_x, -C_z$ and $B_y, C_z \mapsto -B_y, -C_z$. These can be used to map the probability $P_{A|X}(+|0)$ to $P_{A|X}(-|0)$ and the probability $P_{AB|XY}(++|00)$ to any of the probabilities $P_{AB|XY}(+-|00)$, $P_{AB|XY}(-+|00)$, and $P_{AB|XY}(--|00)$, and vice versa. Consequently, in order to derive upper bounds on $G(A|0)$ and $G(AB|00)$, we need only derive concave upper bounds for

$$P_{A|X}(+|0) = \frac{1}{2}\big(1 + \langle A_1 \rangle\big) \tag{5.23}$$

and

$$P_{AB|XY}(++|00) = \frac{1}{4}\big(1 + \langle A_1\rangle + \langle B_1\rangle + \langle A_1 B_1\rangle\big)\,. \tag{5.24}$$

5.3.2 Local guessing probability linearisation

Similarly to the derivation for CHSH summarised above, the local-guessing-probability bound (5.13) for $\sqrt{M^2 + M'^2} \geq 2\sqrt{2}$ is implied by the linearisation

$$\cos(\theta)\langle A_1\rangle + \frac{1}{2}\sin(\theta)\big(\cos(\varphi)M + \sin(\varphi)M'\big) \leq 1 + \sin(\theta)\,, \tag{5.25}$$

which holds for θ in the range $\pi/4 \leq \theta \leq \pi/2$ and for all φ. We can see that (5.25) is tight by observing that is attained if (for example) the measurements

$$A_0 = B_0 = \sigma_{\mathrm{x}}\,, \qquad\qquad A_1 = B_1 = \sigma_{\mathrm{y}} \tag{5.26}$$

and

$$C_0 = \cos(\varphi)\sigma_{\mathrm{x}} - \sin(\varphi)\sigma_{\mathrm{y}},\ C_1 = \sin(\varphi)\sigma_{\mathrm{x}} + \cos(\varphi)\sigma_{\mathrm{y}} \tag{5.27}$$

are performed on the state

$$|\Psi\rangle = \cos(\tfrac{\theta}{2})\frac{1}{\sqrt{2}}\Big(|{+}{+}{+}\rangle + |{+}{-}{-}\rangle\Big) + \sin(\tfrac{\theta}{2})\frac{1}{\sqrt{2}}\Big(|{-}{+}{-}\rangle + |{-}{-}{+}\rangle\Big)\,, \tag{5.28}$$

where $|\pm\rangle = (|1\rangle \pm |2\rangle)/\sqrt{2}$ are the eigenstates of the σ_{x} operator and θ is the same angle as in (5.25). With this state and measurements, one can readily verify that $\langle A_0\rangle = \cos(\theta)$ and $\frac{1}{2}\big(\cos(\varphi)M + \sin(\varphi)M'\big) = 1 + \sin(\theta)$, which attain (5.25).

The linearisation (5.25) ceases to apply for $\theta < \pi/4$. It is violated, for instance, by measuring

$$A_0 = \mathbb{1},\ A_1 = -\mathbb{1},\ B_0 = \sigma_{\mathrm{x}},\ B_1 = \sigma_{\mathrm{y}}\,, \tag{5.29}$$

and

$$C_0 = \cos(\varphi)\sigma_{\mathrm{x}} - \sin(\varphi)\sigma_{\mathrm{y}},\ C_1 = \sin(\varphi)\sigma_{\mathrm{x}} + \cos(\varphi)\sigma_{\mathrm{y}} \tag{5.30}$$

on a state $|\Psi'\rangle = |\chi\rangle_{\mathrm{A}}\,|\psi\rangle_{\mathrm{BC}}$, where $|\chi\rangle$ is any state on Alice's subsystem and Bob and Charlie share the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(e^{-i\frac{\pi}{8}}\,|11\rangle + e^{i\frac{\pi}{8}}\,|22\rangle\Big)\,. \tag{5.31}$$

This strategy yields $\langle A_0 \rangle = 1$ and

$$\cos(\varphi)M + \sin(\varphi)M' = 2\sqrt{2}\,, \tag{5.32}$$

and the right-hand side of (5.25) attains $\cos(\theta) + \sqrt{2}\sin(\theta)$. Importantly, for $\theta = \pi/4$, we see that (5.25) can be attained with a strategy for which Alice's A_0 measurement produces a deterministic outcome.

We prove the linearisation (5.25) by showing that the operator

$$T = \big(1 + \sin(\theta)\big)\mathbb{1} - \cos(\theta)A_0 - \frac{1}{2}\sin(\theta)\big(\cos(\varphi)\hat{M} + \sin(\varphi)\hat{M}'\big) \tag{5.33}$$

is positive semidefinite, where

$$\hat{M} = A_0B_0C_0 - A_0B_1C_1 - A_1B_0C_1 - A_1B_1C_0\,, \tag{5.34}$$
$$\hat{M}' = A_0B_0C_1 + A_0B_1C_0 + A_1B_0C_0 - A_1B_1C_1\,. \tag{5.35}$$

A sum-of-squares decomposition that shows this is

$$T = |P_1^+|^2 + |P_2^+|^2 + |P_1^-|^2 + |P_2^-|^2 \tag{5.36}$$

where $|O|^2 = O^\dagger O$,

$$P_1^+ = \alpha R_1^+ + \beta R_2^+ - \beta R_3^+ - \alpha R_4^+\,, \tag{5.37}$$
$$P_2^+ = \gamma R_1^+ - \delta R_3^+\,, \tag{5.38}$$

$$P_1^- = \beta R_1^- + \alpha R_2^- + \alpha R_3^- + \beta R_4^-\,, \tag{5.39}$$
$$P_2^- = \delta R_1^- + \gamma R_3^-\,, \tag{5.40}$$

$R_i^\pm$ are the operators

$$R_1^+ = \cos(\varphi)(B_0 + C_0) + \sin(\varphi)(B_1 + C_1) - A_0(B_0 + C_0), \tag{5.41}$$
$$R_2^+ = \cos(\theta)(B_1 + C_1) - A_0(B_1 + C_1) + \sin(\theta)A_1(B_0 + C_0), \tag{5.42}$$
$$R_3^+ = \sin(\varphi)(B_0 + C_0) - \cos(\varphi)(B_1 + C_1) - A_0(B_1 + C_1), \tag{5.43}$$
$$R_4^+ = \cos(\theta)(B_0 + C_0) - A_0(B_0 + C_0) - \sin(\theta)A_1(B_1 + C_1), \tag{5.44}$$

$$R_1^- = \cos(\varphi)(B_0 - C_0) + \sin(\varphi)(B_1 - C_1) + A_0(B_0 - C_0), \tag{5.45}$$
$$R_2^- = \cos(\theta)(B_1 - C_1) - A_0(B_1 - C_1) + \sin(\theta)A_1(B_0 - C_0), \tag{5.46}$$
$$R_3^- = \sin(\varphi)(B_0 - C_0) - \cos(\varphi)(B_1 - C_1) + A_0(B_1 - C_1), \tag{5.47}$$
$$R_4^- = \cos(\theta)(B_0 - C_0) - A_0(B_0 - C_0) - \sin(\theta)A_1(B_1 - C_1), \tag{5.48}$$

and the coefficients α, β, γ, δ are related to θ and φ by

$$\alpha = \frac{\sin\left(\frac{\varphi}{2}\right)}{4\cos\left(\frac{\theta}{2}\right)}\,, \tag{5.49}$$

$$\beta = \frac{\cos\left(\frac{\varphi}{2}\right)}{4\cos\left(\frac{\theta}{2}\right)}\,, \tag{5.50}$$

$$\gamma = \frac{1}{4}\sqrt{\sin(\theta) + \cos(\theta)\cos(\varphi) - \sin(\varphi)\sqrt{-\cos(2\theta)}}\,, \tag{5.51}$$

$$\delta = \frac{s}{4}\sqrt{\sin(\theta) - \cos(\theta)\cos(\varphi) + \sin(\varphi)\sqrt{-\cos(2\theta)}}\,, \tag{5.52}$$

where $s = \pm 1$ in the last line is the sign

$$s = -\operatorname{sign}\Big(\cos(\theta)\sin(\varphi) + \cos(\varphi)\sqrt{-\cos(2\theta)}\Big)\,. \tag{5.53}$$

The $R_i^{\pm}$s have been grouped by whether or not they change sign under the replacements

$$\tau\colon \begin{cases} B_0 \mapsto C_0 \\ B_1 \mapsto C_1 \\ C_0 \mapsto B_0 \\ C_1 \mapsto B_1 \end{cases}\,, \tag{5.54}$$

which is a symmetry of (5.33).

The parameters γ and δ are chosen to solve the simultaneous equations

$$8\gamma^2 + 8\delta^2 - \sin(\theta) = 0\,, \tag{5.55}$$

$$8\cos(\varphi)\gamma^2 - 16\sin(\varphi)\gamma\delta - 8\cos(\varphi)\delta^2 - \cos(\theta) = 0\,, \tag{5.56}$$

which we encountered when searching for a decomposition. They are solvable for real-valued γ and δ (and (5.51) and (5.52) are solutions) if $\sin(\theta)$ is positive and greater than $|\cos(\theta)|$, which is the case for the range $\pi/4 \leq \theta \leq \pi/2$ of values of θ for which we need to show that the linearisation (5.25) holds. It is not difficult to check in this case that

$$\sin(\theta) - |\cos(\theta)\cos(\varphi)| \geq |\sin(\varphi)|\sqrt{-\cos(2\theta)} \tag{5.57}$$

holds for arbitrary φ, verifying that the expressions under the outer square roots in (5.51) and (5.52) are nonnegative. The operators $P_i^{\pm}$ are then all Hermitian and $|P_i^{\pm}|^2$ can be simplified to $P_i^{\pm\,2}$.

The Python script `pa1_mermin_sos.py` supplied in [WBA18b] uses the SymPy library [MSP+17] to verify symbolically that the sum-of-squares decomposition (5.36) expands to (5.33), under the assumption that the operators $P_i^{\pm}$ are Hermitian and that the conditions (5.55) and (5.56) for γ and δ can be satisfied.

5.3.3 Two-party guessing probability linearisation

For $M \geq 2$, the guessing-probability bound (5.16) follows from the linearisation

$$\beta\langle A_0 + B_0 + A_0B_0\rangle + \alpha M \leq \gamma\,, \tag{5.58}$$

where

$$\beta = (\lambda - \mu)(\lambda + 3\mu)\,, \tag{5.59}$$
$$\alpha = 4\lambda\mu\,, \tag{5.60}$$
$$\gamma = (3\lambda + \mu)(\lambda + 3\mu)\,, \tag{5.61}$$

which holds for parameters λ and μ satisfying

$$3\mu \geq \lambda \geq \mu\,. \tag{5.62}$$

In the extreme cases $\lambda = 3\mu$ and $\lambda = \mu$, (5.58) reduces respectively to

$$4P_{AB|XY}(++|00) + M \leq 6\,, \tag{5.63}$$

which corresponds to the linear part of (5.16), and to the bound $M \leq 4$ for the Mermin correlator itself, where the gradient of (5.16) is infinite. (Eq. (5.58) also appears to hold for $0 \leq \lambda < \mu$; however (5.58) then translates to a lower bound on $P_{AB|XY}(++|00)$, which we did not interest ourselves in.) Eq. (5.58) is attained with equality by measuring $A_0 = B_0 = C_0 = \sigma_{\mathrm{x}}$ and $A_1 = B_1 = C_1 = \sigma_{\mathrm{y}}$ on the state

$$|\Psi\rangle = \lambda\,|{+}{+}{+}\rangle + \mu\big(|{+}{-}{-}\rangle + |{-}{+}{-}\rangle + |{-}{-}{+}\rangle\big)\,, \tag{5.64}$$

with λ and μ scaled to satisfy $\lambda^2 + 3\mu^2 = 1$ so that the state is properly normalised. In this case $P_{AB|XY}(++|00)$ and M work out to

$$P_{AB|XY}(++|00) = \lambda^2\,, \tag{5.65}$$

and

$$M = (\lambda + 3\mu)^2\,; \tag{5.66}$$

these are related by

$$P_{AB|XY}(++|00) = \frac{3}{4} - \frac{M}{8} + \sqrt{3}\sqrt{\frac{M}{8}\Big(\frac{1}{2} - \frac{M}{8}\Big)}\,, \tag{5.67}$$

corresponding to the nonlinear part of (5.16). The condition $3\mu \geq \lambda \geq \mu$ and normalisation $\lambda^2 + 3\mu^2 = 1$ also translate to precisely the ranges $1/4 \leq P(++|00) \leq 3/4$ and $3 \leq M \leq 4$ to which the nonlinear part of (5.16) applies.

With the same state (5.64) and optimal measurements, we also have

$$P_{ABC|XYZ}(+++|000) = \lambda^2 = P_{AB|XY}(++|00). \tag{5.68}$$

This implies that the upper bound (5.16) for $G(AB|00)$ is also the tight upper bound for $G(ABC|000)$.

The linearisation (5.58) is equivalent to the operator inequality

$$T = \gamma\mathbb{1} - \beta(A_0 + B_0 + A_0B_0) - \alpha\hat{M} \geq 0\,. \tag{5.69}$$

This is shown by the sum-of-squares decomposition

$$\begin{aligned} T = {} & \big|P_1^{++}\big|^2 + \big|P_2^{++}\big|^2 + \big|P_3^{++}\big|^2 + \big|P_4^{++}\big|^2 + \big|P_2^{+-}\big|^2 \\ & + \big|P_1^{-+}\big|^2 + \big|P_2^{-+}\big|^2 + \big|P_1^{--}\big|^2 + \big|P_3^{--}\big|^2\,, \end{aligned} \tag{5.70}$$

where

$$P_1^{++} = \frac{\sqrt{\lambda+\mu}}{4\sqrt{\mu}}(3\mu - \lambda)R_1^{++}\,, \tag{5.71}$$

$$P_2^{++} = \frac{1}{4\sqrt{\mu}}\sqrt{(\lambda^2-\mu^2)(3\mu-\lambda)}\big(R_1^{++} + 2R_2^{++}\big)\,, \tag{5.72}$$

$$P_3^{++} = \frac{\sqrt{3\mu-\lambda}}{2\sqrt{\mu}(\lambda+\mu)}R_3^{++}\,, \tag{5.73}$$

$$P_4^{++} = \frac{1}{2\sqrt{\lambda\mu}}\Big(\frac{\lambda-\mu}{\lambda+\mu}R_3^{++} + R_4^{++}\Big)\,, \tag{5.74}$$

$$P_2^{+-} = \frac{1}{2}\sqrt{\lambda(\lambda-\mu)}R_2^{+-}\,, \tag{5.75}$$

$$P_1^{-+} = \frac{1}{2}\frac{\sqrt{\lambda}}{\sqrt{2\mu}}\sqrt{(\lambda-\mu)^2+4\mu^2}R_1^{-+}\,, \tag{5.76}$$

$$P_2^{-+} = \frac{1}{2}\sqrt{\frac{\lambda(3\mu-\lambda)}{\mu(\lambda+\mu)}}R_2^{-+}\,, \tag{5.77}$$

$$P_1^{--} = \sqrt{\frac{\lambda(\lambda-\mu)}{2}} R_1^{--}\,, \tag{5.78}$$

$$P_3^{--} = \frac{\sqrt{\lambda(\lambda-\mu)}}{2(\lambda+\mu)} \left(R_2^{--} + R_3^{--}\right), \tag{5.79}$$

and

$$R_1^{++} = (A_0 + B_0)(\mathbb{1} - C_0), \tag{5.80}$$

$$R_2^{++} = C_0 - A_0 B_0, \tag{5.81}$$

$$R_3^{++} = (\lambda-\mu)^2 \mathbb{1} + (\lambda+\mu)^2 C_0 - \ (\lambda^2-\mu^2)(A_0+B_0) + 4\lambda\mu\, A_1 B_1, \tag{5.82}$$

$$R_4^{++} = (\lambda-\mu)^2 \mathbb{1} + \mu(\lambda+\mu)(A_0+B_0) - (\lambda^2-\mu^2) C_0 + 2\lambda\mu\,(A_1+B_1) C_1, \tag{5.83}$$

$$R_1^{+-} = (\lambda-\mu)^2 (A_1+B_1) - 2(\lambda-\mu)^2 C_1 - (\lambda^2-\mu^2)(A_1+B_1) C_0 + \ (\lambda^2-\mu^2)(A_0+B_0) C_1, \tag{5.84}$$

$$R_2^{+-} = (A_1+B_1) - 2C_1 - (A_0 B_1 + A_1 B_0) + \ (A_0+B_0) C_1, \tag{5.85}$$

$$R_1^{-+} = (A_0 - B_0)(\mathbb{1} + C_0), \tag{5.86}$$

$$R_2^{-+} = (\lambda+\mu)(A_0 - B_0) - 2\mu(A_1 - B_1) C_1, \tag{5.87}$$

$$R_1^{--} = (A_1 - B_1)(\mathbb{1} - C_0), \tag{5.88}$$

$$R_2^{--} = 2\mu(A_1 - B_1) - (\lambda+\mu)(A_0 - B_0) C_1, \tag{5.89}$$

$$R_3^{--} = (\lambda-\mu)(A_1 - B_1) + \ (\lambda+\mu)(A_0 B_1 - A_1 B_0). \tag{5.90}$$

The $R_i^{\pm\pm\prime}$s are grouped according to whether they change sign under the replacements

$$\tau_1 \colon \begin{cases} A_0 \mapsto B_0 \\ A_1 \mapsto B_1 \\ B_0 \mapsto A_0 \\ B_1 \mapsto A_1 \end{cases}, \qquad \tau_2 \colon \begin{cases} A_1 \mapsto -A_1 \\ B_1 \mapsto -B_1 \\ C_1 \mapsto -C_1 \end{cases}. \tag{5.91}$$

Note that we have included an operator, R_1^{+-}, among the list of $R_i^{\pm\pm\prime}$s that we attempted to construct a sum-of-squares decomposition out of, although ultimately we did not use it.

The Python script `pa1b1_mermin_sos.py`, supplied in [WBA18b], checks that the sum-of-squares decomposition (5.70) expands to (5.69).

5.3.4 Method

We initially determined the upper bounds on the guessing probabilities $G(A|0)$ and $G(AB|00)$ numerically in terms of the Mermin expectation value M. It was quickly apparent that the nonlinear parts of the bounds were consistently being attained with anticommuting measurements. From there it was not difficult to guess the optimal states and see that the numeric bounds seemed to coincide with the (at this point, conjectured) analytic forms (5.13) and (5.16) given in section 5.2. Experimenting a little, we found that the bounds seemed to be attained respectively at the NPA hierarchy levels $1 + \mathrm{AB} + \mathrm{AC}$ and $1 + \mathrm{AB} + \mathrm{AC} + \mathrm{BC}$; this told us that we should be able to find sum-of-squares decompositions out of the operators at these levels for the tangents of the bounds.

We searched for sum-of-squares decompositions following a method similar to [BP15]. The idea is essentially to write the general form of a candidate sum-of-squares decomposition in terms of unknown parameters, assert that it should expand to the operator we want to show is positive semidefinite, and then find parameters for which the assertion becomes true.

Using the tangents of the local-guessing-probability bound as an example, we were searching for a solution to the problem

$$T - \sum_i {P_i^s}^2 = 0\,, \tag{5.92}$$

where T is the target expansion (5.33), for operators $P_i^{\pm}$ of the form

$$P_i^s = \sum_j c_{ij}^s R_j^s\,, \tag{5.93}$$

where the c_{ij}^ss are unknown real-valued coefficients and the R_j^ss form a basis of the space of linear combinations of the operators at level $1 + \mathrm{AB} + \mathrm{AC}$ with the property

$$R_j^s \,|\Psi\rangle = 0 \tag{5.94}$$

for the (conjectured) optimal measurements A_x, B_y, C_z and state $|\Psi\rangle$ described in subsection 5.3.2. Such a basis of R_j^ss is given by Eqs. (5.41)–(5.48).

We have applied some simplifications to the problem above, following [BP15]. In particular, writing

$$\sum_i P_i^\dagger P_i = \sum_{jkrs} M_{jk}^{rs}\, {R_j^r}^\dagger R_k^s\,, \qquad\qquad M_{jk}^{rs} = \sum_i {c_j^r}^*\, c_k^s \tag{5.95}$$

for the potentially more general problem with

$$P_i = \sum_{js} c_{ij}^s R_j^s \,, \tag{5.96}$$

we have used that it is not restrictive to assume that the coefficients c_{ij}^s are real-valued and that the symmetry of the target operator (5.33) under the transformation τ (5.54) can be used to block diagonalise the matrix of elements M_{jk}^{rs}.

We also applied another simplification: one can choose to set $c_{ij}^s = 0$ for (for instance) $i < j$ or $i > j$. This corresponds to choosing a Cholesky factorisation of the matrix of elements $M_{jk}^s = \sum_i c_{ij}^s c_{ik}^s$.

Expanding the candidate sum-of-squares decomposition on the left-hand side of (5.92) and requiring operator-by-operator that the left-hand side is zero translates to imposing a number of quadratic equality constraints on the coefficients c_{ij}^s. We used a Python module `divars.py`, supplied in [WBA18b], together with SymPy, to automate this procedure and help simplify the resulting constraints. We then repeatedly searched numerically for solutions to the constraints, guessing and gradually introducing constraints on the coefficients (e.g., trying $c_{ij}^s = 0$ for some coefficient or imposing that two coefficients are equal to each other) until the numeric search seemed to consistently return the same solution. Solving the remaining constraints by hand got us the sum-of-squares decomposition given in subsection 5.3.2.

5.4 Attacks against device-independent secret sharing

Aside from fundamental interest, a second more practical motivation to conduct the previous analysis was to construct a device-independent secret-sharing protocol based on the Mermin inequality. However, we found obstacles to this idea which we describe in the following section.

5.4.1 Overview

Secret sharing is a cryptographic task in which a secret (e.g., a cryptographic key) is distributed among two or more parties in such a way that a specified minimum number of parties must work together in order to reconstruct it. Hillery, Bužek, and Berthiaume (HBB) [HBcvB99] proposed a quantum version

of secret sharing, analogous to the concept of quantum key distribution, in which the security of the protocol is guaranteed by quantum physics. In the three-party scheme of [HBcvB99], Alice, Bob, and Charlie share a GHZ state $|\Psi\rangle = \frac{1}{\sqrt{2}}\big(|111\rangle + |222\rangle\big)$ and choose inputs $x, y, z \in \{0, 1\}$ and measure A_x, B_y, and C_z, where $A_0 = B_0 = C_0 = \sigma_{\mathrm{x}}$ and $A_1 = B_1 = C_1 = \sigma_{\mathrm{y}}$. In all cases, Bob's and Charlie's measurement outcomes individually are uncorrelated with Alice's. However, if Alice, Bob, and Charlie all measure σ_{x}, or any one of them measures σ_{x} and the other two measure σ_{y}, then Bob and Charlie can together determine Alice's result from the product of their own measurement results. Quantum secret-sharing protocols can also be devised for more than three parties, but we will discuss explicitly only the three-party version here.

The state and measurements, and resulting correlations, of this protocol are precisely those that maximally violate the Mermin-Bell inequality. For readers familiar with both, it may seem natural to ask whether the security of the HBB scheme can be proved device independently, i.e., without assuming that the participants' devices are necessarily measuring σ_{x} and σ_{y}. There have indeed been proposals to design a device-independent secret-sharing protocol based on the GHZ-paradox or other correlations arising from GHZ states [AGCA12, GZ17, RM17]. However, we found that the HBB scheme is completely insecure from a device-independent point of view. The reason is that the secret-sharing protocol is intended to still work, securely, if either Bob or Charlie (but not both) are dishonest, and this differs from the usual Bell scenario where all the parties participating in the Bell test are trusted.

If (say) Charlie is dishonest, he could attack the protocol in the usual ways considered in the security analyses of device-independent protocols (particularly, he could prepare a different state than the GHZ state and/or arrange for Alice's and Bob's devices to perform different measurements than σ_{x} and σ_{y}). Moreover, since Charlie is also involved in the parameter estimation (e.g., the estimation of the Mermin expectation value), he could also act in ways that don't respect the normal conditions of a Bell test:

1. Charlie could wait until Bob declares which basis y he measured in before declaring his own input z and output c, and could perform different measurements on his system depending on which input y Bob declared.

2. Charlie could introduce correlations between his choice of input z and the system prepared for the protocol, instead of choosing z randomly and independently, for instance by performing a four-outcome measurement to determine both his input z and output c, or by implementing a hidden-variable model in which the hidden variable λ is correlated with z.

3. Charlie could perform a different measurement to attempt to guess Alice's outcome than he does in the parameter estimation rounds.

The possibility of an attack combining 1 and 3 is already known to be fatal for even the device-dependent HBB scheme (i.e., Charlie can learn Alice's outcome, without being detected, even assuming that Alice and Bob are measuring σ_{x} and σ_{y}). It and a possible remedy, in which Bob and Charlie are required to declare their outputs before either are allowed to declare their inputs, is discussed in [KKI99].

In the following we describe how a dishonest party could go about attacking an HBB-type protocol, in either the quantum or no-signalling scenarios, without needing to learn Bob's input. We have not attempted to be exhaustive or general; we merely describe the simplest pathological cases that would need to be ruled out, which already show that the situation is much worse for secret sharing in the device-independent scenario.

5.4.2 Hidden variable models

Similarly to other device-independent cryptographic protocols, the simplest way a dishonest Charlie could try to attack a secret-sharing protocol would be to attempt to implement a deterministic hidden-variable model replicating the observed correlations. This is possible if the probabilities $P(abc|xyz)$ of the protocol can be expressed in the form

$$P(abc|xyz) = \sum_{\lambda} p_{\lambda|z} P(a|x,\lambda) P(b|y,\lambda) P(c|z,\lambda). \tag{5.97}$$

Note that, in this case, there is no reason for Charlie to arrange for the hidden variable λ and his own input z to be uncorrelated. (In the language of Bell locality, the so-called "free will" assumption is not justified.) We reflect this in (5.97) by allowing the probability distribution $p_{\lambda|z}$ to depend arbitrarily on z. Eq. (5.97) thus does not have the form of a local hidden-variable model of the kind normally considered in Bell-type theorems, and it is not sufficient for the probabilities $P(abc|xyz)$ to violate a Bell inequality, such as the Mermin inequality, in order to rule out a local hidden-variable model of the form above.

It is easy to show that the existence of a decomposition of the form (5.97) is equivalent to the existence of a local hidden-variable model of the form

$$P(ab|xy,cz) = \sum_{\lambda} p'_{\lambda|cz} P^{(cz)}(a|x,\lambda) P^{(cz)}(b|y,\lambda) \tag{5.98}$$

for each of the probability distributions $P(ab|xy, cz)$ conditioned on Charlie's different possible outputs and inputs c and z. This gives a bare minimum condition in order for there to be any hope that a device-independent secret-sharing scheme might be secure: at least one of the conditional distributions $P(ab|xy, cz)$ (for some c and z) must be nonlocal. This condition is not met for the GHZ correlations that the HBB protocol is based on: in that case all of the conditional distributions $P(ab|xy; cz)$ exhibit perfect correlation or no correlation at all depending on the inputs, and admit trivial local hidden-variable models. This makes it clear that secret sharing cannot be done securely and device independently using only the correlations of the GHZ paradox.

5.4.3 No-signalling attacks

Security analyses of device-independent protocols are sometimes undertaken using only the no-signalling constraints, since this is typically much simpler, though typically at the cost of significantly worse tolerance to noise. We are aware of at least two proposals [AGCA12, GZ17] to design device-independent secret-sharing protocols using GHZ states (but not necessarily the GHZ-paradox correlations) using only no-signalling constraints. In this case, the situation is significantly worse, since in the no-signalling scenario, a dishonest Charlie could implement arbitrary steering. More precisely, suppose Charlie wishes to produce the no-signalling distribution $P(abc|xyz)$ in the parameter estimation rounds. If the marginal distribution $P(ab|xy) = \sum_c P(abc|xyz)$ can be expressed as a convex sum

$$P(ab|xy) = \sum_{\lambda} p_\lambda P^{(\lambda)}(ab|xy) \tag{5.99}$$

of no-signalling distributions $P^{(\lambda)}(ab|xy)$ then Charlie could prepare the extended distribution

$$P'(abc|xyz) = \begin{cases} P(abc|xyz) & \text{if } z \neq \perp \\ p_c P^{(c)}(ab|xy) & \text{if } z = \perp \end{cases} \tag{5.100}$$

where $\perp$ is an additional input that Charlie can use in the secret bit generation rounds, when he is not asked to publicly disclose his input and outcome. It is easy to verify that the extended distribution (5.100) still satisfies the no-signalling constraints.

The above observation means that, in the no-signalling scenario, the security or insecurity of a device-independent secret-sharing protocol against a dishonest Charlie is determined entirely by the marginal distribution $P(ab|xy)$ between Alice and Bob. If this marginal distribution is in the local polytope then the

protocol is completely insecure against no-signalling attacks. A special case worth remarking is that no device-independent secret-sharing protocol based on the GHZ state can be proved secure using only the no-signalling conditions: the marginals of the GHZ state are all separable and the marginal probability distributions will always be in the local polytope, regardless of what or how many measurements are performed by the parties.

5.4.4 Outlook

We have pointed out that a device-independent version of the HBB protocol would be completely insecure against a dishonest party, and that any protocol for which the marginal probability distributions are in the local polytope (for example, any protocol using a GHZ state) cannot be proved secure using only the no-signalling constraints. This does not rule out that a device-independent secret-sharing protocol could be designed, for instance based on different correlations and/or using stronger constraints than only the no-signalling conditions in the security proof. However, one should consider the following points:

- It is already known that if one can do quantum key distribution then one can do secret sharing. For instance, Alice could do device-independent key distribution separately with Bob and Charlie and xor the two keys. More generally, secret sharing can be done securely using classical protocols if the parties can do one-time-pad encryption, which happens to be precisely what key distribution schemes are intended to generate cryptographic keys for.

- As with key distribution, or any secure protocol involving parties communicating remotely, the parties would need to authenticate themselves. This is normally done in key distribution using classical authentication schemes which require preshared keys; part of the generated key can then be used to do the authentication the next time. Consequently, it seems to us that one would need to be able to do key distribution anyway in order to do secret sharing, if only to generate the authentication keys needed after the first use of the protocol.

Given these issues, the usefulness of a device-independent secret-sharing protocol that does not reduce to a direct application of device-independent quantum key distribution is unclear to us.

5.5 Conclusion

We considered the Mermin-Bell experiment with three parties and we identified and proved tight upper bounds on the guessing probabilities associated with the measurement outcomes of one and two of the parties. The results are fundamental tradeoffs between the amount of intrinsic randomness and nonlocality, as measured by the violation of the Mermin inequality, imposed by the structure of quantum physics. The linearisations in section 5.3 can also be read as inequalities identifying parts of the boundary of the set of quantum correlations. The results reveal that part of the boundary of the quantum set is flat, a characteristic that has previously been remarked upon in [RM17, GKW$^+$18].

It may be interesting to study how our results generalise to Bell experiments involving more parties. We guessed one possible generalisation of the upper bound (5.16) for $G(AB|00)$ to n parties, which can be found in appendix C.2. We did not attempt to prove it, though we tested the cases for $n = 4$ and $n = 5$ parties numerically.

While we are not aware of an obvious practical application of our results, we believe there is some merit to finding the analytic form of randomness vs. nonlocality tradeoffs more generally where it could be feasible to do so, particularly where the result might be used in the security proof of a device-independent protocol. From this point of view, our results explore the feasibility of searching for sum-of-squares decompositions for problems somewhat larger than was considered in [BP15]. The cases where the method is likely to work are probably those where the problem is "simple" in some same key respects as the problems we studied. In particular: it was reasonably easy for us to guess the upper bounds and the states and measurements that attained them, we found that the optimal solution was attained at a level of the hierarchy that was not prohibitively high, and symmetries of the problem allowed us to reduce the number of variables in the searches for sum-of-squares decompositions.

Chapter 6

Two bits of global randomness from any partially entangled state

When two parties perform a Bell test with dichotomic measurements, up to two bits of global randomness can be generated. The amount of randomness that is produced can be certified in a device-independent way, that is, without relying on the physical system in use, but only on the correlations that underlie the Bell test. Though, in that case, the state on which the measurements are performed does not play a role in the randomness certification process, one might take a resource approach to entanglement and ask: how much randomness can be certified when the underlying correlations arise from a state with a given level of entanglement? Indeed, at first glance, the notions of entanglement and randomness are closely related: randomness can be certified only from nonlocal correlations, which, in turn, can be achieved only with entangled states. However, we show in this Chapter that this relation is only qualitative: maximal randomness can be certified from any level of entanglement. The results of this Chapter are based on [WKB+19].

6.1 Introduction

Randomness as we understand it in this Thesis, that is, the unpredictability of the outcomes of a Bell test, exists only if the behaviour associated to the Bell test is nonlocal. In turn, nonlocal correlations can be observed only if measurements are performed on an entangled sate. This might suggest that there exists a quantitative equivalence between the notions of randomness and entanglement.

Aside from it fundamental interest, that would imply some experimental requirements on the states that should be prepared in order to certify a certain amount of randomness in a device-independent way. Finding out if it such requirements exist is crucial, as practical and efficient DIRG is now within our reach [LYL+18, BKG+18].

However, the authors of [AMP12] already observed that correlations that arise from almost unentangled states can be used to certify an amount of randomness that is arbitrarily close to its maximal value. In this Chapter, we go a step forward and ask: can maximal global randomness be certified with correlations that arise from any partially entangled qubit pure state?

We show that the answer is yes. To do so, we fix the values of four Bell expressions. We prove that these values self-test the desired partially entangled qubit pure state, and the measurements that are needed to obtain two bits of global randomness.

6.2 Results

We first describe the setting that we consider. We then prove our self-testing claim. We conclude by explaining why it implies that maximal global randomness is certified.

6.2.1 Setting

An arbitrary two-qubit pure state can be expressed as

$$|\psi_\theta\rangle = \cos\left(\tfrac{\theta}{2}\right)|00\rangle + \sin\left(\tfrac{\theta}{2}\right)|11\rangle \tag{6.1}$$

in its Schmidt decomposition. In a device-dependent approach, if Alice and Bob share such a state, they can extract two bits of randomness from it by measuring, for instance, $A = \sigma_{\mathrm{x}}$ on Alice's side and $B = \sigma_{\mathrm{y}}$ on Bob's side. This is the maximum amount of randomness that can be extracted from $|\psi_\theta\rangle$ using

measurements that are projective on the support of Alice's and Bob's marginals of $|\psi_\theta\rangle$.

We now describe the scheme we use to achieve the same thing in a device-independent way. Alice performs three projective measurements A_0, A_1, A_2 and Bob performs seven projective measurements $B_0, \ldots, B_6$ on a state ρ that is a priori unknown but intended to be $|\psi_\theta\rangle$ in some basis for some $\theta \in]0, \frac{\pi}{2}]$. These measurements are dichotomic and we denote their possible outputs $\{+1, -1\}$. They check that the correlations they obtain satisfy the conditions

$$I_\beta = 2\sqrt{2}\sqrt{1+\beta^2/4}, \tag{6.2}$$
$$J_\beta = 2\sqrt{2}\sqrt{1+\beta^2/4}, \tag{6.3}$$
$$S = 2\sqrt{2}\sin(\theta), \tag{6.4}$$
$$\langle A_2 B_6\rangle = -\sin(\theta), \tag{6.5}$$

where

$$I_\beta = \beta\langle A_0\rangle + \langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \langle A_1 B_1\rangle\,, \tag{6.6}$$
$$J_\beta = \beta\langle A_0\rangle + \langle A_0 B_2\rangle + \langle A_0 B_3\rangle + \langle A_2 B_2\rangle - \langle A_2 B_3\rangle \tag{6.7}$$

are tilted CHSH expressions of the kind introduced in [AMP12], we choose

$$\beta = \frac{2\cos(\theta)}{\sqrt{1+\sin(\theta)^2}}\,, \tag{6.8}$$

and

$$S = \langle A_1 B_4\rangle + \langle A_1 B_5\rangle + \langle A_2 B_4\rangle - \langle A_2 B_5\rangle \tag{6.9}$$

is the ordinary CHSH expression.

Before we state our exact claims, let us describe the general reasoning behind these requirements. It was shown in [AMP12] that the tilted CHSH expressions I_β and J_β have a maximum quantum expectation value of $2\sqrt{2}\sqrt{1+\beta^2/4}$. Furthermore, this quantum bound is attained with the partially entangled state $|\psi_\theta\rangle$, with θ related to β according to (6.8), and measurements $A_0 = \sigma_{\mathrm{z}}$ and A_1 and A_2 in the X-Y plane on Alice's side. Given this, the third condition $S = 2\sqrt{2}\sin(\theta)$ can only be satisfied if A_1 and A_2 are orthogonal on the bloch sphere, i.e., we have something like $A_1 = \sigma_{\mathrm{x}}$ and $A_2 = \sigma_{\mathrm{y}}$. The final condition $\langle A_2 B_6\rangle = -\sin(\theta)$ would then require $B_6 = \sigma_{\mathrm{y}}$. At this point, Alice and Bob could trust that they can extract two bits of global randomness with the measurements $A_1 = \sigma_{\mathrm{x}}$ and $B_6 = \sigma_{\mathrm{y}}$.

6.2.2 Self-testing

The density operator ψ_θ associated to the state $|\psi_\theta\rangle$ can be written as:

$$\psi_\theta = \frac{1}{4}\Big[\mathbb{1}\otimes\mathbb{1}+\cos(\theta)\big(\mathbb{1}\otimes\sigma_{\mathrm{z}}+\sigma_{\mathrm{z}}\otimes\mathbb{1}\big)+\sin(\theta)\big(\sigma_{\mathrm{x}}\otimes\sigma_{\mathrm{x}}-\sigma_{\mathrm{y}}\otimes\sigma_{\mathrm{y}}\big)+\sigma_{\mathrm{z}}\otimes\sigma_{\mathrm{z}}\Big]. \quad (6.10)$$

If the first condition $I_\beta = 2\sqrt{2}\sqrt{1+\beta^2/4}$ is met, we can infer that, in a suitable choice of basis, the underlying quantum state has the form

$$\rho = \psi_\theta \otimes \sigma_{\mathrm{junk}} \quad (6.11)$$

where σ_{junk} is unknown and the measurements A_0 and A_1 on Alice's side are

$$A_0 = \sigma_{\mathrm{z}} \otimes \mathbb{1}, \quad (6.12)$$
$$A_1 = \sigma_{\mathrm{x}} \otimes \mathbb{1}. \quad (6.13)$$

This self-testing statement is proved in detail in Appendix D.

The second condition $J_\beta = 2\sqrt{2}\sqrt{1+\beta^2/4}$ allows us to make an analogous self-testing claim for the state and measurements A_0 and A_2, although not necessarily in the same basis. However, from the first condition, we have already determined the state and A_0, and the self test tells us that A_2 must be related to A_0 by

$$\{A_0,\, A_2\} = 0 \quad (6.14)$$

regardless of the choice of basis. Writing generally

$$A_2 = \mathbb{1} \otimes A_{\mathbb{1}} + \sigma_{\mathrm{x}} \otimes A_{\mathrm{X}} + \sigma_{\mathrm{y}} \otimes A_{\mathrm{Y}} + \sigma_{\mathrm{z}} \otimes A_{\mathrm{Z}}, \quad (6.15)$$

imposing (6.14) with $A_0 = \sigma_{\mathrm{z}} \otimes \mathbb{1}$ forces $A_{\mathbb{1}} = A_{\mathrm{Z}} = 0$. Requiring in addition that ${A_2}^2 = \mathbb{1} \otimes \mathbb{1}$, we find that the measurement A_2 must have the form

$$A_2 = \sigma_{\mathrm{x}} \otimes A_{\mathrm{X}} + \sigma_{\mathrm{y}} \otimes A_{\mathrm{Y}} \quad (6.16)$$

with

$${A_{\mathrm{X}}}^2 + {A_{\mathrm{Y}}}^2 = \mathbb{1}, \quad [A_{\mathrm{X}},\, A_{\mathrm{Y}}] = 0. \quad (6.17)$$

We now study the third constraint $S = 2\sqrt{2}\sin(\theta)$ and show that it implies $A_{\mathrm{X}} = 0$. Writing

$$B_i = \mathbb{1} \otimes B_{\mathbb{1}}^{(i)} + \mathrm{X} \otimes B_{\mathrm{X}}^{(i)} + \mathrm{Y} \otimes B_{\mathrm{Y}}^{(i)} + \mathrm{Z} \otimes B_{\mathrm{Z}}^{(i)}, \quad (6.18)$$

$i \in \{4,5\}$, the condition ${B_i}^2 = \mathbb{1} \otimes \mathbb{1}$ implies:

$$ {B_{\mathbb{1}}^{(i)}}^2 + {B_{\mathrm{X}}^{(i)}}^2 + {B_{\mathrm{Y}}^{(i)}}^2 + {B_{\mathrm{Z}}^{(i)}}^2 = \mathbb{1}. \tag{6.19} $$

That implies in particular:

$$ {B_{\mathrm{X}}^{(i)}}^2 + {B_{\mathrm{Y}}^{(i)}}^2 \leq \mathbb{1}. \tag{6.20} $$

We now express S with these notations. Using the expression (6.10) for ψ_θ in the Pauli basis and the fact that the Pauli operators are traceless, we get:

$$ \begin{aligned} \langle A_2 B_4 \rangle &= \mathrm{Tr}\big[A_2 B_4 \left(\psi_\theta \otimes \sigma_{\mathrm{junk}}\right)\big] \\ &= \sin(\theta)\big(\langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(4)} \rangle_{\mathrm{junk}} - \langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(4)} \rangle_{\mathrm{junk}}\big) \end{aligned} \tag{6.21} $$

and, similarly:

$$ \langle A_2 B_5 \rangle = \sin(\theta)\left(\langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(5)} \rangle - \langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(5)} \rangle\right), \tag{6.22} $$

$$ \langle A_1 B_4 \rangle = \sin(\theta)\,\langle \mathbb{1} \otimes B_{\mathrm{X}}^{(4)} \rangle, \tag{6.23} $$

$$ \langle A_1 B_5 \rangle = \sin(\theta)\,\langle \mathbb{1} \otimes B_{\mathrm{X}}^{(5)} \rangle. \tag{6.24} $$

The condition $S = 2\sqrt{2}\sin(\theta)$ thus translates to:

$$ \begin{aligned} &\langle \mathbb{1} \otimes B_{\mathrm{X}}^{(4)} \rangle + \langle \mathbb{1} \otimes B_{\mathrm{X}}^{(5)} \rangle + \langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(4)} \rangle \\ &- \langle A_{\mathrm{Y}} \otimes B_{\mathrm{X}}^{(4)} \rangle - \langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(5)} \rangle + \langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(5)} \rangle = 2\sqrt{2}. \end{aligned} \tag{6.25} $$

Since $\mathbb{1}, A_{\mathrm{X}}$ and A_{Y} commute, we can co-diagonalise them. Together with the fact that ${A_{\mathrm{X}}}^2 + {A_{\mathrm{Y}}}^2 = \mathbb{1}$, we can thus write:

$$ \mathbb{1} = \sum_k |k\rangle\langle k|, \qquad A_{\mathrm{X}} = \sum_k x_k |k\rangle\langle k|, \qquad A_{\mathrm{Y}} = \sum_k y_k |k\rangle\langle k| \tag{6.26} $$

with

$$ \forall k,\ {x_k}^2 + {y_k}^2 = 1. \tag{6.27} $$

Using this, we have, for instance,

$$ \begin{aligned} \langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(4)} \rangle &= \sum_k x_k\, \mathrm{Tr}\big[\big(|k\rangle\langle k| \otimes B_{\mathrm{X}}^{(4)}\big)\sigma_{\mathrm{junk}}\big] \\ &= \sum_k x_k \langle B_{\mathrm{X}}^{(4)} \rangle_k \end{aligned} \tag{6.28} $$

and similar expressions for the other terms on the left side of (6.25), where the expectation values $\langle \cdot \rangle_k = \mathrm{Tr}[\,\cdot\,\sigma_k]$ are evaluated on the states

$$\sigma_k = \mathrm{Tr}_{\mathrm{A_{junk}}}\big[\big(|k\rangle\langle k| \otimes \mathbb{1}\big)\sigma_{\mathrm{junk}}\big] \tag{6.29}$$

on the 'junk' part of the Hilbert space on Bob's side. Note that their norms satisfy

$$\sum_k \|\sigma_k\|^2 = \sum_k \mathrm{Tr}[\sigma_k] = 1\,. \tag{6.30}$$

Using this, together with the Cauchy-Schwarz inequality, on the left-hand side of (6.25), we get:

$$\begin{aligned}
&\langle \mathbb{1} \otimes (B_{\mathrm{X}}^{(4)} + B_{\mathrm{X}}^{(5)})\rangle + \langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(4)}\rangle - \langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(4)}\rangle - \langle A_{\mathrm{X}} \otimes B_{\mathrm{X}}^{(5)}\rangle + \langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(5)}\rangle \\
&\quad = \sum_k \Big[(1+x_k)\langle B_{\mathrm{X}}^{(4)}\rangle_k - y_k\langle B_{\mathrm{Y}}^{(4)}\rangle_k + (1-x_k)\langle B_{\mathrm{X}}^{(5)}\rangle_k + y_k\langle B_{\mathrm{Y}}^{(5)}\rangle_k\Big] \\
&\quad \leq \sum_k \Big(\sqrt{2(1+x_k)}\sqrt{\langle B_{\mathrm{X}}^{(5)}\rangle_k{}^2 + \langle B_{\mathrm{Y}}^{(5)}\rangle_k{}^2} + \sqrt{2(1-x_k)}\sqrt{\langle B_{\mathrm{X}}^{(6)}\rangle_k{}^2 + \langle B_{\mathrm{Y}}^{(6)}\rangle_k{}^2}\Big) \\
&\quad \leq \sum_k \Big(\sqrt{2(1+x_k)}\,\|\sigma_k\|^2 + \sqrt{2(1-x_k)}\,\|\sigma_k\|^2\Big) \\
&\quad \leq \sum_k 2\sqrt{2}\,\|\sigma_k\|^2 \\
&\quad = 2\sqrt{2}.
\end{aligned} \tag{6.31}$$

The third expression comes from Eq. (6.27), the fourth expression comes from the fact that

$$\langle B\rangle_k \leq \sqrt{\langle B^2\rangle_k}\,\|\sigma_k\| \tag{6.32}$$

and

$$\big\langle B_{\mathrm{X}}^{(i)}{}^2 + B_{\mathrm{Y}}^{(i)}{}^2\big\rangle_k \leq \langle \mathbb{1}\rangle_k = \|\sigma_k\|^2. \tag{6.33}$$

Finally, for Eq. (6.25) to hold, all the inequalities in (6.31) should actually be equalities. In particular, the vectors:

$$\begin{pmatrix}\sqrt{2(1+x_k)}\\ \sqrt{2(1-x_k)}\end{pmatrix} \qquad \begin{pmatrix}\|\sigma_k\|^2\\ \|\sigma_k\|^2\end{pmatrix} \tag{6.34}$$

should be collinear, which is only possible if $x_k = 0$ for all k. In other words, $A_{\mathrm{X}} = 0$, and we conclude that Alice's third measurement must be of the form

$$A_2 = \sigma_{\mathrm{y}} \otimes A_{\mathrm{Y}} \tag{6.35}$$

with ${A_{\mathrm{Y}}}^2 = \mathbb{1}$.

Let us now write B_6 in the same way as we did for B_4 and B_5:

$$B_6 = \mathbb{1} \otimes B_{\mathbb{1}}^{(6)} + \sigma_{\mathrm{x}} \otimes B_{\mathrm{X}}^{(6)} + \sigma_{\mathrm{y}} \otimes B_{\mathrm{Y}}^{(6)} + \sigma_{\mathrm{z}} \otimes B_{\mathrm{Z}}^{(6)} \tag{6.36}$$

with

$${B_{\mathbb{1}}^{(6)}}^2 + {B_{\mathrm{X}}^{(6)}}^2 + {B_{\mathrm{Y}}^{(6)}}^2 + {B_{\mathrm{Z}}^{(6)}}^2 = \mathbb{1}\,. \tag{6.37}$$

We now compute $|\langle A_2 B_6\rangle|$:

$$\begin{aligned} |\langle A_2 B_6\rangle| &= \left|\langle \sigma_{\mathrm{y}} \otimes \sigma_{\mathrm{y}}\rangle_{\psi_\theta} \langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(6)}\rangle_{\mathrm{junk}}\right| \\ &= |\sin(\theta)|\left|\langle A_{\mathrm{Y}} \otimes B_{\mathrm{Y}}^{(6)}\rangle\right| \\ &\leq |\sin(\theta)|\sqrt{\langle {B_{\mathrm{Y}}^{(6)}}^2\rangle}\sqrt{\langle {A_{\mathrm{Y}}}^2\rangle} \\ &= |\sin(\theta)|\sqrt{\langle {B_{\mathrm{Y}}^{(6)}}^2\rangle}. \end{aligned} \tag{6.38}$$

The condition $\langle A_2 B_6\rangle = -\sin(\theta)$ then allows us to conclude that

$$\left\langle {B_{\mathrm{Y}}^{(6)}}^2 \right\rangle = 1 \tag{6.39}$$

and, from (6.37), that

$$\left\langle {B_{\mathbb{1}}^{(6)}}^2 \right\rangle = \left\langle {B_{\mathrm{X}}^{(6)}}^2 \right\rangle = \left\langle {B_{\mathrm{Z}}^{(6)}}^2 \right\rangle = 0. \tag{6.40}$$

6.2.3 Maximal randomness certification

We can now show that the probabilities of the possible outcomes when Alice and Bob jointly measure A_1 and B_6 are all 1/4. Indeed, given Eqs. (6.11) and (6.13), the following holds:

$$\begin{aligned} |\langle A_1\rangle| &= |\langle \sigma_{\mathrm{x}} \otimes \mathbb{1}\rangle_{\psi_\theta} \langle \mathbb{1} \otimes \mathbb{1}\rangle_{\mathrm{junk}}| \\ &= 0. \end{aligned} \tag{6.41}$$

Moreover, Eq. (6.40) implies that:

$$\begin{aligned} |\langle B_6\rangle| &= \left|\langle \mathbb{1} \otimes \mathbb{1}\rangle_{\psi_\theta} \langle \mathbb{1} \otimes B_{\mathbb{1}}^{(6)}\rangle_{\mathrm{junk}} + \langle \mathbb{1} \otimes \sigma_{\mathrm{z}}\rangle_{\psi_\theta} \langle \mathbb{1} \otimes B_{\mathrm{Z}}^{(6)}\rangle_{\mathrm{junk}}\right| \\ &\leq \left|\langle \mathbb{1} \otimes B_{\mathbb{1}}^{(6)}\rangle\right| + \cos(\theta)\left|\langle \mathbb{1} \otimes B_{\mathrm{Z}}^{(6)}\rangle\right| \\ &\leq \sqrt{\left\langle {B_{\mathbb{1}}^{(6)}}^2 \right\rangle} + \cos(\theta)\sqrt{\left\langle {B_{\mathrm{Z}}^{(6)}}^2 \right\rangle} \\ &= 0. \end{aligned} \tag{6.42}$$

and that:

$$\begin{aligned} |\langle A_1 B_6 \rangle| &= \left| \langle \sigma_{\mathrm{x}} \otimes \sigma_{\mathrm{x}} \rangle_{\psi_\theta} \langle \mathbb{1} \otimes B_{\mathrm{X}}^{(6)} \rangle_{\mathrm{junk}} \right| \\ &= |\sin(\theta)| \left| \langle \mathbb{1} \otimes B_{\mathrm{X}}^{(6)} \rangle \right| \\ &\leq |\sin(\theta)| \sqrt{\langle {B_X^{(6)}}^2 \rangle} \\ &= 0. \end{aligned} \tag{6.43}$$

We thus find, for all quantum realisations compatible with the four conditions (6.2)–(6.5), that:

$$P(ab|x=1,y=6) \leq \frac{1}{4}\Big(1 + |\langle A_1 \rangle| + |\langle B_6 \rangle| + |\langle A_1 B_6 \rangle|\Big) = \frac{1}{4}. \tag{6.44}$$

This implies that $G(AB|1,6) = 1/4$, i.e., two bits of global randomness are certified when performing measurements A_1 and B_6.

Let us add that, if Alice and Bob are not limited to projective measurements, but can perform POVM, they could in principle certify up to two bits of local randomness and four bits of global randomness. The authors of [APVW16] presented a construction that achieves the maximal local value. One can combine their argument and the proof presented in this Chapter to show that this value can be moreover achieved with any partially entangled qubit state.

6.3 Conclusion

When two agents perform two dichotomic measurements, they can generate up to two bits of global randomness. In that case, the CHSH inequality does not certify that much randomness, but other schemes do (see [MP13], or [DPA13] together with [ŠASA16]).

Since randomness can be certified in a device-independent way only if the measurements are performed on an entangled state, one could think that there exists a quantitative connection between randomness and entanglement. One way to address this question is to study whether one has to impose constraints on the entanglement of the underlying state in order to certify maximal randomness. The authors of [AMP12] already proved that almost maximal randomness could be certified with an almost unentangled state. We here proved that the exact maximal value of 2 bits of global randomness can be certified from any partially entangled qubit state.

This result also suggests that practical DIRG with high rates could be achieved without demanding resources, namely, maximally entangled state. However, our scheme requires three measurements for Alice, seven for Bob, and the observation of four Bell expression values, which could be experimentally challenging to implement. One might then look for construction simpler than ours, or might couple our construction with other results where DIRG is also studied from a resource point of view [BMP18].

Chapter 7

Conclusions and outlook

The random nature of quantum theory is perplexing. This property was long debated by the founding fathers of quantum mechanics, and several of them, chief among them Albert Einstein, tried to refute it. However, the fact that quantum theory gives rise to nonlocal correlations proves that it cannot be completed in a way that would make its laws deterministic. In turn, nonlocality-based randomness is guaranteed to be intrinsic: the fact that the outcomes of a Bell test are unpredictable cannot be attributed to a lack of knowledge. Deriving trade-offs between nonlocality and randomness, as we did in this Thesis, has two objectives. First, it allows us to understand quantum theory in terms of its (un)predictive power. Second, it gives us the ability to certify, in a device-independent way, that a Bell test generates private random bits, a crucial resource for numerous cryptographic protocols.

In this Chapter, we remind the different frameworks in which we derived such trade-offs. We then briefly state the results that we obtained, and their implications. Finally, we describe possible future lines of investigation for these different frameworks.

Randomness based on no-signalling and time-ordering We have looked into how much randomness can be certified when one does not rely on the quantum formalism, but only on the no-signalling principle. In that case, the formalism is simpler, but also poorer. At the moment, it is not clear whether, in that simple framework, one could obtain a sequence of perfect private random bits of any desired length by using a Bell test. Previous investigations have

explored whether it is possible to create a sequence of perfect private random bits from a longer sequence of partially private random bits, that is, whether privacy amplification is possible. In this Thesis, we investigated a prior problem: can a sequence of Bell tests generate a sequence of partially private random bits of any length? If the Bell tests are assumed to be causally independent, the answer is yes. If no causal structure is assumed, the answer is no. But the question remains open if one makes the natural assumption of time-ordering, i.e., if one assumes that past Bell tests have an influence on future Bell tests, but not the other way around. We proved that, in that case, the unpredictability of the outputs of each Bell test decreases with each repetition, but not as much as when no causal structure is imposed. This shows that time-ordering entails a fundamental difference for randomness, compared with the two above-mentioned cases, even in the case where only no-signalling is assumed.

To answer the question in the negative, one would have to show that the min-entropy rate of the outcomes of all the Bell tests, and even its smooth version, tends to zero as the number of repetition increases. To answer in the positive, one would have to show that the min-entropy rate tends to infinity when the number of repetitions increases. Our results and the framework we developed might help in proving one of these two statements.

Practical randomness generation via regularisation A black-box approach to quantum information aims to derive relations between mathematical objects such as a behaviour and information-related quantities such as randomness: they hold for a theoretical Bell test and its associated underlying behaviour. However, in a practical Bell experiment, the behaviour is not accessible, as it is an ideal mathematical object. If one directly infers the underlying behaviour from the frequencies collected in a real Bell experiment, one obtains a behaviour that does not obey the no-signalling principle, due to finite statistics. In order to evaluate a device-independent quantity on that behaviour, it should then be first regularised, that is, projected onto the no-signalling or quantum set. In this Thesis, we investigated how well this approach works for certifying randomness. More precisely, we compared two approaches to DIRG: the first one gives a bound on the min-entropy of the Bell tests' outputs based on a pre-determined witness of nonlocality; the second one requires to first estimate and regularise the underlying behaviour, to then derive a better suited nonlocality witness for that specific behaviour, that is, one that yields a better randomness versus nonlocality trade-off. We carried out several numerical simulations that show that the second approach is more favourable, in the sense that it certifies more

randomness.

Our method relies on a derivation of min-entropy bounds that holds only against an eavesdropper with classical-side information. This restriction is reasonable for DIRG, as no quantum information needs to be sent on a public channel. A possible future line of investigation would be to extend it to the case of quantum-side information. It could then be used for DIQKD, where it might yield higher key rates. However, the current derivations of min-entropy bounds against quantum-side information heavily relies on specific nonlocality witnesses. One would thus need to derive such bounds that hold for generic witnesses.

Randomness in the tripartite scenario As nonlocality was primarily defined on correlations between two agents, so was nonlocality-based randomness. However, the concept of nonlocality can be extended to more than two parties. In this Thesis, we derived analytical trade-offs between randomness and nonlocality for three parties, that is, we evaluated how much randomness could be certified from the outcomes of one, two, and three parties, when these three parties evaluate the violation of a tripartite Bell inequality, namely, the Mermin inequality. The original motivation was to use such trade-offs to design device-independent secret sharing protocols. However, we provided strong arguments that tend to show that secret sharing is not compatible with the device-independent approach. One reason for that is that the evaluation of a Bell inequality violation is meaningful only if the different parties collaborate and trust each other, an assumption that one cannot make in the framework of secret sharing. Yet the results that we obtained are valuable on their own, as the trade-offs that we obtained are surprising: the generalisation of the guessing probability to many parties seems to be far from trivial.

Finding applications of such trade-offs to other multi-partite information protocols would enables us to perform new tasks in a device-independent way. However, even if they might not have such applications, deriving similar trade-offs for other Bell inequalities, or for more parties, could help us understand the specificities of quantum correlations through the prism of randomness.

Maximal randomness from partial entanglement At first glance, the notions of entanglement and nonlocality-based randomness are deeply correlated. Indeed, performing measurements on a quantum state can give rise to nonlocal correlations only if the state is entangled. However, it was already observed that almost maximal randomness could be certified even when the underlying state is almost not entangled. The maximal amount of global randomness that

can be certified when two parties perform dichotomic measurements is 2 bits. In this Thesis, we proved that this maximal amount could be certified in a device-independent way from any partially entangled pure qubit state. To do so, we showed that the observation of some specific values for four Bell expressions provides a self-test for such partially entangled state and measurements that yield this maximal amount of randomness. This proves that the relation between randomness and entanglement is only qualitative.

We proved that claim with ten measurements and four Bell expressions. There is no evidence that this requirement is minimal, and one might prove that it holds with less measurements or less Bell expressions. Moreover, we studied only two-dimensional systems. One could investigate the link between nonlocality-based randomness and entanglement in systems of higher dimension.

Appendices

A. Appendix of Chapter 3

We prove the claims of Chapter 3. In Appendix A.1, we prove the guessing probability value for $n = 1$ (Eq. 3.2). We then prove that G_n is independent of the choice of inputs $(\mathbf{x}^*, \mathbf{y}^*)$ (App. A.2) and that the product of n perfect PR-correlations is a vertex of the polytopes associated with any of the no-signalling type constraints introduced in this work (App. A.3).

We proceed with the proofs of all the guessing probability values given in Table 3.1. We first simplify the optimization problem (3.14) using symmetry arguments, and we give its general expression, as well as its associated dual formulation, that follow from these symmetries (App. A.4). The detailed expression of the feasible points for these two problems that yield the same objective function value can be found in [BPA18a]. We provide the necessary information about how to read these files in Appendix A.5.

A.1 Solution for $n = 1$

We first give a feasible point for (3.14) that attains the bound given in Eq. (3.2). Let $\{D_i\}_{i=1}^4$ be four deterministic behaviors defined as:

$$\begin{aligned} D_1(a,b|x,y) &= \delta_{a,0}\delta_{b,0}, \\ D_2(a,b|x,y) &= \delta_{a,x}\delta_{b,0}, \\ D_3(a,b|x,y) &= \delta_{a,0}\delta_{b,y}, \\ D_4(a,b|x,y) &= \delta_{a,x}\delta_{b,y+1}. \end{aligned} \tag{A.1}$$

Take

$$P(\alpha = 0) = P(\alpha = 1) = \frac{1}{2},$$

$$P^{\alpha=0}(ab|xy) = \frac{1-v}{4}\sum_{i=1}^{4} D_i(ab|xy) + v\mathrm{PR}_1(ab|xy), \tag{A.2}$$

$$P^{\alpha=1}(ab|xy) = P^{\alpha=0}(\overline{ab}|xy),$$

where, for $s \in \{0,1\}$, $\overline{s}$ denotes its complement. Then $\{P(\alpha), P^{\alpha}\}_{\alpha\in\{0,1\}}$ is a feasible point for (3.14) that has objective value $1-\frac{v}{2}$. Moreover, when $n=1$, Eq. (3.25) implies $G_1(v) \leq 1-\frac{v}{2}$. This concludes the proof of Eq. (3.2).

A.2 Symmetries of the guessing probability problem

The following transformations allow us to express (3.14) in a reduced form.

Lemma 1. *Let $(T_1^i), (T_2^i)$ and (T_3^i) be transformations that map a behaviour $P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})$ onto another behaviour by re-ordering its inputs and outputs in the following way:*

$$(T_1^i): \begin{cases} a_i \to \overline{a_i} \\ b_i \to \overline{b_i} \\ \alpha_i \to \overline{\alpha_i} \end{cases}, \qquad (T_2^i): \begin{cases} a_i \to a_i \oplus x_i \\ y_i \to \overline{y_i} \end{cases}, \qquad (T_3^i): \begin{cases} b_i \to b_i \oplus y_i \\ x_i \to \overline{x_i} \end{cases}.$$

Then, for all i and for all the NS *conditions, $(T_1^i), (T_2^i)$ and (T_3^i) map a feasible point for* (3.14) *onto another feasible point. Moreover, (T_1^i) preserves the objective function value for all possible* NS *conditions, and (T_2^i) preserves the objective function value for* {Full-NS, ABNS, TONS}.

Proof. We first prove that a feasible point is mapped onto another feasible point. For a given round i, let (T_j^i) be one these transformations and let $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ be a feasible point for (3.14) for some NS condition. Let $\{\tilde{P}(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ be the image of this point by (T_j^i). Since the NS condition involves all $(\mathbf{a},\mathbf{b},\mathbf{x},\mathbf{y})$, and since (T_j^i) simply reorders some elements of $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ in an individual round i, $\{\tilde{P}(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ satisfies the same NS condition. Moreover, since the behavior $\mathrm{PR}_v(a_ib_i|x_iy_i)$ is invariant under (T_j^i), $\{\tilde{P}(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ also satisfies the constraint on the marginals. (T_j^i) thus maps a feasible point for (3.14) onto another feasible point.

We now show that (T_1^i) preserves the objective function value of (3.14) for all the NS conditions. For simplicity, let us take $i=1$, the argument for $i>1$

being the same. Let $\{\tilde{P}(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ be the image of a feasible point $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ by (T_1^1). Then:

$$\begin{aligned}\sum_{\boldsymbol{\alpha},\mathbf{b}} \tilde{P}(\boldsymbol{\alpha})\tilde{P}^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},\mathbf{0}) &= \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\overline{\alpha_1}\boldsymbol{\alpha}_{>\mathbf{1}})P^{\overline{\alpha_1}\boldsymbol{\alpha}_{>1}}(\overline{\alpha_1}\boldsymbol{\alpha}_{>\mathbf{1}}, \overline{b_1}\mathbf{b}_{>\mathbf{1}}|\mathbf{0},\mathbf{0}) \\ &= \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},\mathbf{0})\end{aligned} \tag{A.3}$$

We now show that (T_2^i) preserves the objective function value of (3.14) for all but the WTONS condition. We again set $i=1$, and denote $\{\tilde{P}(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ the image by (T_2^1) of a feasible point $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ for the Full-NS, ABNS or TONS condition. Then:

$$\begin{aligned}\sum_{\boldsymbol{\alpha},\mathbf{b}} \tilde{P}(\boldsymbol{\alpha})\tilde{P}^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},\mathbf{0}) &= \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},10\ldots0) \\ &= \sum_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha}) \sum_{\mathbf{b}} P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},10\ldots0) \\ &= \sum_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha}) \sum_{\mathbf{b}} P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},00\ldots0)\end{aligned} \tag{A.4}$$

where the last equality holds because, for all α, $P_{\boldsymbol{\alpha}}$ is ABNS. □

Thanks to (T_2^i) and (T_3^i), we can now prove that the optimal value $G_n(v)$ defined in (3.14) is independent of $(\mathbf{x}^*,\mathbf{y}^*)$, i.e., $G_n(\mathbf{A}|\mathbf{x}^*,\mathbf{y}^*)[v] = G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v]$. Let us assume that $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ is a feasible point for (3.14) that achieves the value $G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v]$. We then construct $\{P(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ by applying (T_2^1) onto $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$. Then:

$$\begin{aligned}\sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})\tilde{P}^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|00\ldots0,10\ldots0) &= \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|00\ldots0,00\ldots0) \\ &= G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v]\end{aligned} \tag{A.5}$$

This implies $G_n(\mathbf{A}|00\ldots0,10\ldots0)[v] \geq G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v]$. Let us now assume that $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ is a feasible point for (3.14) that achieves the value $G_n(\mathbf{A}|00\ldots0,10\ldots0)[v]$, and construct $\{P(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ by applying (T_2^1) onto it. Then:

$$\begin{aligned}\sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})\tilde{P}^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|00\ldots0,00\ldots0) &= \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|00\ldots0,10\ldots0) \\ &= G_n(\mathbf{A}|00\ldots0,10\ldots0)[v]\end{aligned} \tag{A.6}$$

This implies $G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v] \geq G_n(\mathbf{A}|00\ldots0,10\ldots0)[v]$ and thence $G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v] = G_n(\mathbf{A}|00\ldots0,10\ldots0)[v]$. The same construction can be done for all other values of $\mathbf{y}$ by applying (T_2^i) whenever $y_i = 1$, thus proving $G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v] = G_n(\mathbf{A}|\mathbf{0},\mathbf{y})[v]$ for all $\mathbf{y}$.

We now assume that $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ is a feasible point for (3.14) that achieves the value $G_n(\mathbf{A}|\mathbf{0},\mathbf{y})[v]$. For some $\mathbf{x} \in \{0,1\}^n$, we construct $\{P(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ by applying (T_3^i) onto it whenever $x_i = 1$. Then:

$$\begin{aligned}\sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})\tilde{P}^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{x},\mathbf{y}) = \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}\oplus\mathbf{xy}|\mathbf{0},\mathbf{y}) = \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},\mathbf{y}) \\ = G_n(\mathbf{A}|\mathbf{0},\mathbf{y})[v]\end{aligned} \tag{A.7}$$

where the first equality holds because we applied (T_3^i) only when $x_i = 1$ and the second one holds because we sum over $\mathbf{b}$. This implies that $G_n(\mathbf{A}|\mathbf{x},\mathbf{y})[v] \geq G_n(\mathbf{A}|\mathbf{0},\mathbf{y})[v]$. Let us now assume that $\{P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ is a feasible point for (3.14) that achieves the value $G_n(\mathbf{A}|\mathbf{x},\mathbf{y})[v]$. We construct $\{P(\boldsymbol{\alpha}), \tilde{P}^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})\}$ by applying (T_3^i) onto it whenever $x_i = 1$. Then:

$$\begin{aligned}\sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})\tilde{P}^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},\mathbf{y}) = \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}\oplus\mathbf{xy}|\mathbf{x},\mathbf{y}) = \sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha})P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{x},\mathbf{y}) \\ = G_n(\mathbf{A}|\mathbf{x},\mathbf{y})[v]\end{aligned} \tag{A.8}$$

This implies $G_n(\mathbf{A}|\mathbf{0},\mathbf{y})[v] \geq G_n(\mathbf{A}|\mathbf{x},\mathbf{y})[v]$ thence $G_n(\mathbf{A}|\mathbf{x},\mathbf{y})[v] = G_n(\mathbf{A}|\mathbf{0},\mathbf{y})[v]$. Altogether, this proves that $G_n(\mathbf{A}|\mathbf{x},\mathbf{y})[v] = G_n(\mathbf{A}|\mathbf{0},\mathbf{0})[v]$ for all $(\mathbf{x},\mathbf{y})$, and thus that $G_n(v)$ is properly defined.

A.3 Product of n perfect PR-correlations

We now show that the product of n PR-boxes is a vertex of any of the no-signalling polytopes we introduced in Chapter 3. We do it for $n = 2$, the generalisation to $n \geq 3$ is straightforward. Let us assume that there exists two ABNS (resp. WTONS) joint distributions P_1 and P_2 such that:

$$\mathrm{PR}_1(a_1,b_1|x_1,y_1)\times\mathrm{PR}_1(a_2,b_2|x_2,y_2) = \lambda_1 P_1(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) + \lambda_2 P_2(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) \tag{A.9}$$

for some $(\lambda_1,\lambda_2) \in [0,1]$ such that $\lambda_1 + \lambda_2 = 1$.

Then:

$$\begin{aligned}\mathrm{PR}_1(a_1,b_1|x_1,y_1) &= \lambda_1 \sum_{a_2,b_2} P_1(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) + \lambda_2 \sum_{a_2,b_2} P_2(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) \\ &= \lambda_1 P_1(a_1,b_1|x_1,x_2,y_1,y_2) + \lambda_2 P_2(a_1,b_1|x_1,x_2,y_1,y_2).\end{aligned} \tag{A.10}$$

Let us fix a specific value (x_2^*, y_2^*) for (x_2, y_2). Then $P_1(a_1,b_1|x_1,x_2^*,y_1,y_2^*)$ is a no-signalling bipartite binary behaviour. Indeed,

$$\sum_{b_1} P_1(a_1,b_1|x_1,x_2^*,y_1,y_2^*) = \sum_{a_2,b_1,b_2} P_1(a_1,a_2,b_1,b_2|x_1,x_2^*,y_1,y_2^*) \tag{A.11}$$

is independent of y_1 because P_1 is ABNS (resp. WTONS), and, for the same reason, $\sum_{a_1} P_1(a_1,b_1|x_1,x_2^*,y_1,y_2^*)$ is independent of x_1. The same goes for $P_2(a_1,b_1|x_1,x_2^*,y_1,y_2^*)$. Since the PR-box is a vertex of the polytope of bipartite binary no-signalling behaviours, Eq. (A.10) implies

$$P_1(a_1,b_1|x_1,x_2^*,y_1,y_2^*) = P_2(a_1,b_1|x_1,x_2^*,y_1,y_2^*) = \mathrm{PR}_1(a_1,b_1|x_1,y_1) \tag{A.12}$$

for all values of (x_2^*, y_2^*). The same holds for $P_1(a_2,b_2|x_1^*,x_2,y_1^*,y_2)$ for all values of (x_1^*, y_1^*), as well as for $P_2(a_2,b_2|x_1^*,x_2,y_1^*,y_2)$. That implies:

$$P_1(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) = P_2(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) = \mathrm{PR}_1(a_1,b_1|x_1,y_1) \times \mathrm{PR}_1(a_2,b_2|x_2,y_2). \tag{A.13}$$

The product of two PR-boxes cannot be decomposed over different joint distributions in ABNS (resp. WTONS): it is thus a vertex of the ABNS (resp. WTONS) polytope. Since Full-NS and TONS are subsets of these polytopes, it also implies that it is a vertex of Full-NS and TONS.

A.4 Primal and dual form of the guessing probability problem

The symmetry (T_1^i) given in Appendix A.2 implies that the solutions to the problem defined by Equation (3.14) can be found in the reduced space:

$$\mathcal{S} = \left\{ (P(\boldsymbol{\alpha}), P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})) \middle| P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) = P^{\boldsymbol{\alpha}=\mathbf{0}}(\overline{\mathbf{a},\mathbf{b}}^{\alpha}|\mathbf{x},\mathbf{y}), P(\boldsymbol{\alpha}) = \frac{1}{2^n} \right\} \tag{A.14}$$

where $\overline{a_i,b_i}^{\alpha_i} = \begin{cases} a_i, b_i \text{ if } \alpha_i = 0, \\ \overline{a_i}, \overline{b_i} \text{ if } \alpha_i = 1. \end{cases}$

From here on, we'll thus only consider distributions with such symmetries, and we'll write $P(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})$ for $P_{\boldsymbol{\alpha}=\mathbf{0}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y})$. Note that, for $P \in \mathcal{S}$, the objective function of (3.14) becomes

$$\sum_{\boldsymbol{\alpha},\mathbf{b}} P(\boldsymbol{\alpha}) P^{\boldsymbol{\alpha}}(\boldsymbol{\alpha},\mathbf{b}|\mathbf{0},\mathbf{0}) = \sum_{\boldsymbol{\alpha},\mathbf{b}} \frac{1}{2^n} P^{\mathbf{0}}(\mathbf{0},\mathbf{b}|\mathbf{0},\mathbf{0}) = \sum_{\mathbf{b}} P(\mathbf{0},\mathbf{b}|\mathbf{0},\mathbf{0}). \tag{A.15}$$

Moreover, a constraint on the marginals is now expressed in the following way:

$$\begin{aligned} &\sum_{\boldsymbol{\alpha}} P(\boldsymbol{\alpha}) P^{\boldsymbol{\alpha}}(\mathbf{a},\mathbf{b}|\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}_v(a_i,b_i|x_i,y_i) \\ &\quad \Leftrightarrow \frac{1}{2^n} \sum_{\boldsymbol{\alpha}} P^{\boldsymbol{\alpha}=\mathbf{0}}(\overline{\mathbf{a},\mathbf{b}}^{\alpha}|\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}_v(a_i,b_i|x_i,y_i) \\ &\quad \Leftrightarrow \frac{1}{2^n} \sum_{\boldsymbol{a}} P(\mathbf{a},\mathbf{a}\oplus\mathbf{b}|\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} \mathrm{PR}_v(0,b_i|x_i,y_i). \end{aligned} \tag{A.16}$$

The optimisation problem defined in (3.14) can thus be written as:

$$\begin{aligned} G_n(v) = \quad & \max \quad \sum_{\mathbf{b}} P(\mathbf{0},\mathbf{b}|\mathbf{0},\mathbf{0}) \\ & \text{s.t.} \quad \forall(\mathbf{x},\mathbf{y},\mathbf{b}),\ \sum_{\boldsymbol{a}} P(\mathbf{a},\mathbf{a}\oplus\mathbf{b}|\mathbf{x},\mathbf{y}) = 2^n \times \prod_{i=1}^{n} \mathrm{PR}_v(0,b_i|x_i,y_i) \\ & \quad\ P \in \mathrm{NS} \end{aligned} \tag{A.17}$$

In order to construct the dual of (A.17), note that P can be seen as a vector, on which two kinds of constraints apply: on the one hand positivity, as it represents some probability distributions, on the other hand linear constraints, that arise both from the marginal constraints and the no-signalling scenario that is considered.

The optimisation problem (A.17) and its associated dual problem can then be summarised as:

$$\begin{aligned} G_n(v) = \quad & \max \quad \mathbf{c}^\top \mathbf{p} \\ & \text{s.t.} \quad A\mathbf{p} = \mathbf{b} \\ & \quad\ \mathbf{p} \geq 0 \end{aligned} \tag{A.18}$$

$$\begin{aligned} G_n(v) = \quad & \min \quad \mathbf{b}^\top \mathbf{y} \\ & \text{s.t.} \quad A^\top \mathbf{y} \geq \mathbf{c} \end{aligned} \tag{A.19}$$

where and A and $\mathbf{b}$ describe the marginal and no-signaling constraints and $c_i = \begin{cases} 1 \text{ if } p_i = P(\mathbf{0},\mathbf{b}|\mathbf{0},\mathbf{0}), \\ 0 \text{ otherwise.} \end{cases}$

Strong duality holds here because (A.18) is linear and feasible (the target correlation, i.e., n noisy i.i.d. PR boxes, is always a solution). This implies that finding the optimum now amounts to finding feasible points for these two problems that yield the same objective function value.

A.5 Solutions of the primal and dual problems

The solutions of (A.18) and (A.19) when $n = 2, 3$ can be found in [BPA18b]. Since $G_n(v)$ is the same for TONS and WTONS when $n = 2, 3$, we give only a primal feasible point for TONS and a dual feasible point for WTONS with the same objective function value, which is sufficient to prove the values given in Table 3.1 for TONS and WTONS. Indeed, let us call momentarily p^*_{TONS} (resp. p^*_{WTONS}) the solution of (A.18) for TONS (resp. WTONS), and d^*_{WTONS} the solution of (A.19) for WTONS. Let us call p_{TONS} the objective function value associated to our primal feasible point for TONS, and d_{WTONS} the objective function value associated to our dual feasible point for WTONS. We then have

$$p_{TONS} \leq p^*_{TONS}, \tag{A.20}$$

$$d^*_{WTONS} \leq d_{WTONS}. \tag{A.21}$$

Moreover, $p^*_{TONS} \leq p^*_{WTONS}$ because TONS $\subset$ WTONS and $p^*_{WTONS} = d^*_{WTONS}$ because strong quality holds. Altogether, this gives:

$$p_{TONS} \leq p^*_{TONS} \leq d^*_{WTONS} \leq d_{WTONS}. \tag{A.22}$$

Finding a primal feasible point for TONS and a dual feasible point for WTONS such that $p_{TONS} = d_{WTONS}$ is thus sufficient to solve (3.14) both for TONS and WTONS.

For the solutions of (A.18), we give only $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y} = \mathbf{0})$: since the symmetry (T_2^i) is valid both for TONS and ABNS, the distributions for other values of $\mathbf{y}$ can be derived from $P(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y} = \mathbf{0})$ alone, by applying the corresponding transformation.

For the solutions of (A.19) to be defined without ambiguity, the order of the constraints listed in the matrix A and vector $\mathbf{b}$ should be fixed. We thus include in [BPA18b] the scripts that construct the specific matrices A and vectors $\mathbf{b}$ for which our dual solutions are defined.

B. Appendix of Chapter 4

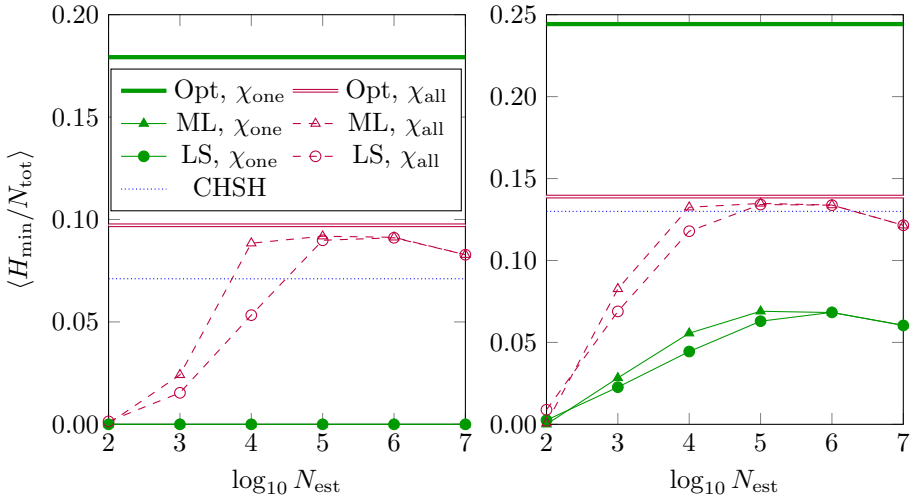
B.1 Tuning the parameters

We present here the analysis that we conducted in order to tune the parameters of the protocol presented in Chapter 4. We first generated four random distributions, in the same way as explained in Chapter 4, and computed the min-entropy rates for varying N_{est}, to see how many bits should be sacrificed for estimation. We fix $\epsilon = \epsilon' = 10^{-6}$, we divide the interval $[I_l, I_q^+]$ in $M + 1 = 1000$ segments of the same length, we use the NPA local level 2 [MBL$^+$13] for the regularisation and the guessing probability problems, we set $N_{\text{tot}} = 10^8$, and we run 500 simulations for each point. We compute the average min-entropy rates $\langle H_{\min}/N_{\text{tot}} \rangle$ as a function of $\log_{10} N_{\text{est}}$ for both regularisation methods ML and LS, and with two possible choices for χ: $\chi_{\text{all}} = \{0,1\}^2$ and $\chi_{\text{one}} = (x^*, y^*)$, where (x^*, y^*) is the most random input pair, i.e. the one that yields the highest RB function. In that case, we set the input distribution to $P_{XY}(x^*, y^*) = \pi_{x^*y^*} = 0.9$ (and uniform on the other inputs). The results are presented in Figure B.1.

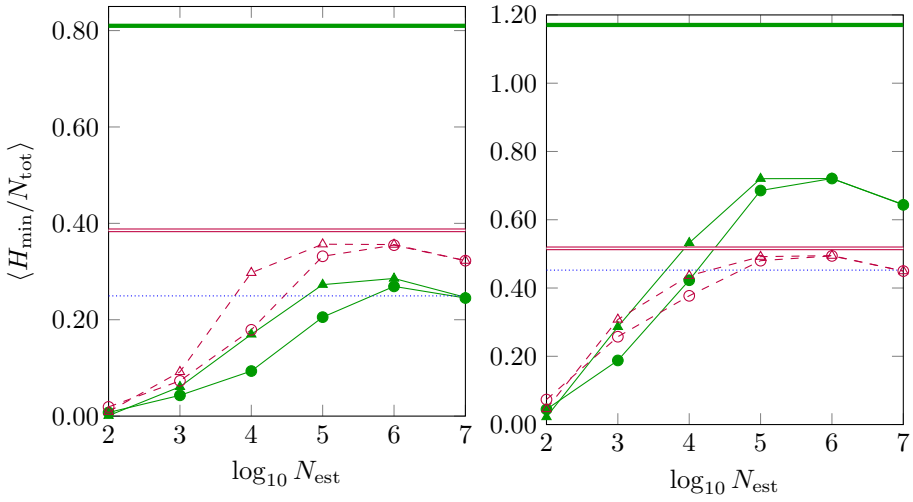
From those graphs, we deduce that setting $N_{\text{est}} = 10^6$, i.e., 1% of the total data, is optimal. Note that, to distinguish these four distributions, we give their CHSH values I_{CHSH}. It does not mean that the CHSH inequality is the best Bell expression for certifying randomness from these behaviours: we merely give it as a way to quantify how nonlocal these distributions are, because it might be interesting for the reader to see that the effects we observe seem to depend on that. For instance, generating randomness from only one input seems to give an advantage only when the CHSH value is high enough.

We then study, under the same conditions, the effect of the input in the bias distribution $\pi_{x^*y^*}$, to see if one can observe an advantage when setting $\chi = \chi_{\text{one}}$ instead of $\chi = \chi_{\text{all}}$. The results can be found in Figure B.2.

We observe that for three distributions, no advantage is obtained when generating randomness from only one input pair, independently of how the input

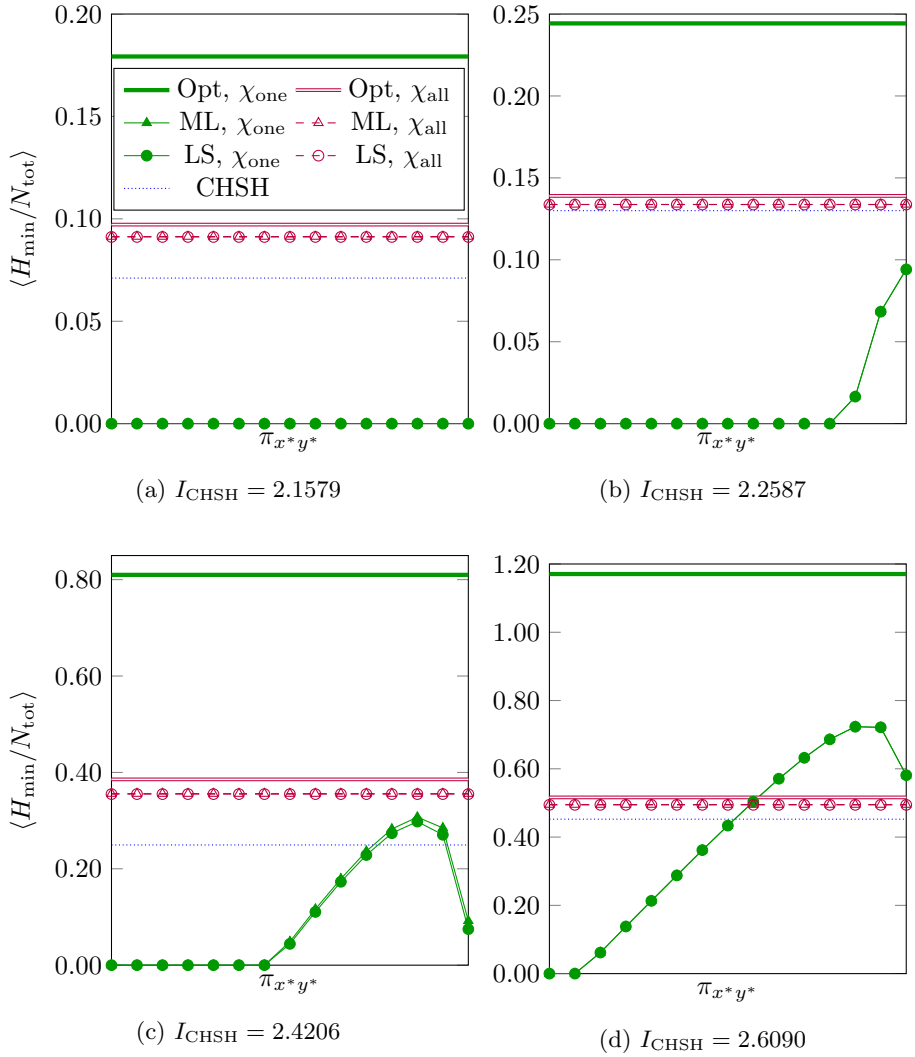


(a) $I_{\text{CHSH}} = 2.1579$ (b) $I_{\text{CHSH}} = 2.2587$



(c) $I_{\text{CHSH}} = 2.4206$ (d) $I_{\text{CHSH}} = 2.6090$

Figure B.1: Average min-entropy rates as a function of the size of the data that is sacrificed for estimation.



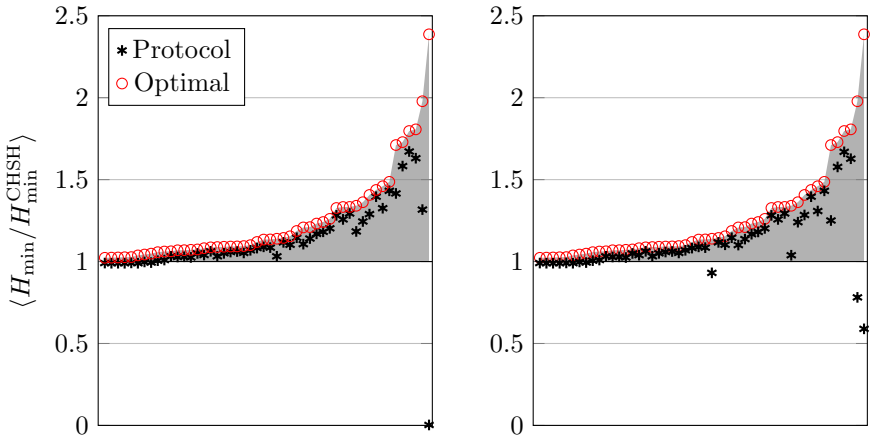
(a) $I_{\text{CHSH}} = 2.1579$

(b) $I_{\text{CHSH}} = 2.2587$

(c) $I_{\text{CHSH}} = 2.4206$

(d) $I_{\text{CHSH}} = 2.6090$

Figure B.2: Average min-entropy rates as a function of the input distribution. In most cases, both regularisation methods give the same value for χ_{one}, which is why they cannot be distinguished.



(a) Regularisation via ML

(b) Regularisation via LS.

Figure B.3: Black asterisk: ratio between the rate obtained via our protocol and via the direct use of the CHSH inequality. Red circle: ratio between the maximal achievable min-entropy and the rate obtained via the direct use of the CHSH inequality.

distribution is biased towards that input pair. That confirms the observation based on the first graph: setting $\chi = \chi_{\mathrm{one}}$ can give an advantage only for the behaviour with highest CHSH value. This is not surprising when one compares these results with the examples provided in [NSBSP18], where the authors also observed that generating randomness from one input pair starts giving an advantage only for high enough $N_{\mathrm{tot}} > 10^8$. We thus decided not to use this possibility and to set $\chi = \chi_{\mathrm{all}}$.

We then compared the min-entropy rates obtained from the ML and LS regularisations. In that case, there is no varying parameter, so we decided to directly run the simulations described in Section 4.3.3 for both regularisations, and to compare the obtained ratios $\langle H_{\min}/H_{\min}^{\mathrm{CHSH}} \rangle$. The results can be found in Figure B.3.

The ML regularisation performs better than the LS regularisation in 98% of the cases. Moreover, while the protocol based on ML performs well for 98% of the cases, that holds for LS only in 94% of the cases. This led us to claim that when one wants to regularise data in order to certify randomness, one should preferably minimise the KL divergence.

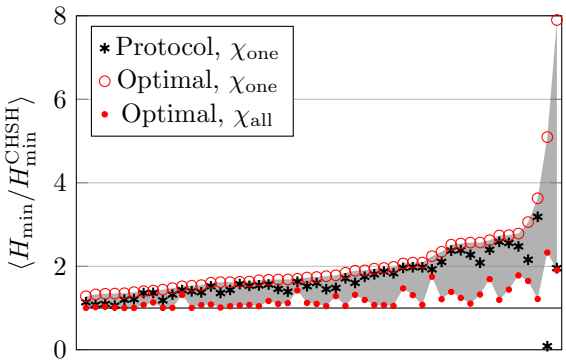


Figure B.4: Asterisk: ratio between the rate obtained via our protocol for χ_{one} and via the direct use of the CHSH inequality. Circle: ratio between the maximal achievable min-entropy for χ_{one} and the rate obtained via the CHSH inequality. Dot: ratio between the maximal achievable min-entropy for χ_{all} and the rate obtained via the CHSH inequality.

B.2 Generating randomness from one input pair

To ensure that our method could result in better min-entropy bounds for $\chi = \chi_{\text{one}}$ when the total number of rounds is big enough, we carried out the same simulations as the ones presented in the main text, but with $N_{\text{tot}} = 10^{12}$. In that case, our method allows us to identify which input pair (x^*, y^*) yields the most favourable RB function, thanks to the ML regularised distribution. We then bias the input distribution towards that pair, setting $\pi_{x^*y^*} = 0.99$. The results are presented in Figure B.4, where we plot the ratios between the min-entropy rate obtained via our protocol and via the direct use of the CHSH inequality $H_{\min}^{\text{CHSH}}$, as well as the ratios between $-\log_2(G^{\chi}_{full}(P_{AB|XY}))$ and $H_{\min}^{\text{CHSH}}$, for $\chi = \chi_{\text{one}}$ and $\chi = \chi_{\text{all}}$. We highlighted in grey the region between these two ratios. 98% of the simulations led to points falling in that region. In those cases, our protocol is good in two ways: not only it performs better than the direct use of CHSH, but it also achieves a higher ratio than the optimal one for all inputs. In that case, the advantage of our protocol is twofold: it allows us to identify the most favourable input pair, and then to taylor the Bell inequality to that specific input pair.

C. Appendix of Chapter 5

We give the no-signalling guessing probability for the Mermin-Bell experiment with three parties presented in Chapter 5. We then give a conjecture on the quantum guessing probability for the same experiment with n parties.

C.1 No-signalling bounds

The main text gave tight bounds on the guessing probability assuming all the measurements are performed on a quantum system. The tightest bound that can be derived for the local guessing probability using only the no-signalling constraints is

$$G(A|0) \leq \frac{3}{2} - \frac{1}{8}|M| - \frac{1}{8}|M'|\,. \tag{C.1}$$

For the two-party guessing probability there are two distinct tight bounds,

$$G(AB|00) \leq \frac{3}{2} - \frac{1}{4}|M| \tag{C.2}$$

and

$$G(AB|00) \leq \frac{7}{4} - \frac{1}{4}|M| - \frac{1}{8}|M'|\,, \tag{C.3}$$

as well as the same bounds with M and M' swapped. Finally, there are three bounds for the global guessing probability,

$$G(ABC|000) \leq \frac{3}{2} - \frac{1}{4}|M|\,, \tag{C.4}$$

$$G(ABC|000) \leq \frac{7}{4} - \frac{1}{4}|M| - \frac{1}{8}|M'|\,, \tag{C.5}$$

$$G(ABC|000) \leq \frac{7}{4} - \frac{1}{16}|M| - \frac{5}{16}|M'|\,. \tag{C.6}$$

The same upper bounds hold for $G(ABC|011)$, $G(ABC|101)$, and $G(ABC|110)$. The upper bounds for $G(ABC|001)$, $G(ABC|010)$, $G(ABC|100)$, and $G(ABC|111)$ are the same except with M and M' swapped.

Following an approach similar to [SPM13], the local-guessing-probability bound (C.1) is implied by the eight inequalities

$$1 - \langle A_0 \rangle + \langle B_0 C_0 \rangle - \langle A_0 B_0 C_0 \rangle \geq 0\,, \tag{C.7}$$
$$1 - \langle A_0 \rangle + \langle B_0 C_1 \rangle - \langle A_0 B_0 C_1 \rangle \geq 0\,, \tag{C.8}$$
$$1 - \langle A_0 \rangle + \langle B_1 C_0 \rangle - \langle A_0 B_1 C_0 \rangle \geq 0\,, \tag{C.9}$$
$$1 - \langle A_0 \rangle - \langle B_1 C_1 \rangle + \langle A_0 B_1 C_1 \rangle \geq 0\,, \tag{C.10}$$
$$1 + \langle A_1 \rangle - \langle B_0 C_0 \rangle - \langle A_1 B_0 C_0 \rangle \geq 0\,, \tag{C.11}$$
$$1 - \langle A_1 \rangle - \langle B_0 C_1 \rangle + \langle A_1 B_0 C_1 \rangle \geq 0\,, \tag{C.12}$$
$$1 - \langle A_1 \rangle - \langle B_1 C_0 \rangle + \langle A_1 B_1 C_0 \rangle \geq 0\,, \tag{C.13}$$
$$1 + \langle A_1 \rangle + \langle B_1 C_1 \rangle + \langle A_1 B_1 C_1 \rangle \geq 0\,. \tag{C.14}$$

Each of these is in turn implied by two positivity constraints. For example, (C.7) is just stating that

$$4P(-{+}{+}|000) + 4P(---|000) \geq 0\,. \tag{C.15}$$

The inequalities (C.7) to (C.14) sum to

$$8 - 4\langle A_0 \rangle - M - M' \geq 0 \tag{C.16}$$

which, together with symmetries of the problem, implies (C.1).

The upper bound $G(AB|00) \leq 3/2 - M/4$ is similarly implied by the five inequalities

$$1 + \langle C_0 \rangle - \langle A_0 B_0 \rangle - \langle A_0 B_0 C_0 \rangle \geq 0\,, \tag{C.17}$$
$$1 - \langle A_0 \rangle - \langle B_1 C_1 \rangle + \langle A_0 B_1 C_1 \rangle \geq 0\,, \tag{C.18}$$
$$1 - \langle B_0 \rangle - \langle A_1 C_1 \rangle + \langle A_1 B_0 C_1 \rangle \geq 0\,, \tag{C.19}$$
$$1 - \langle C_0 \rangle - \langle A_1 B_1 \rangle + \langle A_1 B_1 C_0 \rangle \geq 0\,, \tag{C.20}$$
$$1 + \langle A_1 B_1 \rangle + \langle A_1 C_1 \rangle + \langle B_1 C_1 \rangle \geq 0 \tag{C.21}$$

(the last of these is just stating that

$$4P(+{+}{+}|111) + 4P(---|111) \geq 0\,), \tag{C.22}$$

which sum to

$$6 - 4P_{AB|XY}(++|00) - M \geq 0\,. \tag{C.23}$$

The second upper bound (C.3) for $G(AB|00)$ is implied by the inequalities

$$\begin{aligned}
2 + 2\langle C_0\rangle - 2\langle A_0B_0\rangle - 2\langle A_0B_0C_0\rangle &\geq 0\,, && \text{(C.24)}\\
1 - \langle A_0\rangle + \langle B_0C_1\rangle - \langle A_0B_0C_1\rangle &\geq 0\,, && \text{(C.25)}\\
1 - \langle C_0\rangle + \langle A_0B_1\rangle - \langle A_0B_1C_0\rangle &\geq 0\,, && \text{(C.26)}\\
1 - \langle B_0\rangle + \langle A_1C_0\rangle - \langle A_1B_0C_0\rangle &\geq 0\,, && \text{(C.27)}\\
2 - \langle A_0\rangle - \langle C_1\rangle - \langle A_0B_1\rangle & \\
- \langle B_1C_1\rangle + 2\langle A_0B_1C_1\rangle &\geq 0\,, && \text{(C.28)}\\
2 - \langle A_1\rangle - \langle B_0\rangle - \langle A_1C_1\rangle & \\
- \langle B_0C_1\rangle + 2\langle A_1B_0C_1\rangle &\geq 0\,, && \text{(C.29)}\\
2 - \langle B_1\rangle - \langle C_0\rangle - \langle A_1B_1\rangle & \\
- \langle A_1C_0\rangle + 2\langle A_1B_1C_0\rangle &\geq 0\,, && \text{(C.30)}\\
1 + \langle A_1\rangle + \langle B_1\rangle + \langle C_1\rangle + \langle A_1B_1\rangle & \\
+ \langle A_1C_1\rangle + \langle B_1C_1\rangle + \langle A_1B_1C_1\rangle &\geq 0\,, && \text{(C.31)}
\end{aligned}$$

which sum to

$$14 - 8P_{AB|XY}(++|00) - 2M - M' \geq 0\,. \tag{C.32}$$

Each of the eight inequalities above can be obtained from up to three positivity constraints. For instance, the left-hand side of (C.28) is equal to

$$4P(+--|011) + 8P(-+-|011) + 4P(--+|011)\,. \tag{C.33}$$

The first two upper bounds (C.4) and (C.5) on the three-outcome guessing probability $G(ABC|000)$ are implied by (C.2) and (C.3). Using symmetries of the problem, the remaining inequality (C.6) reduces to showing that

$$\max\bigl(P(+++|000), P(---|111)\bigr) \leq \frac{7}{4} - \frac{1}{16}M - \frac{1}{16}M'. \tag{C.34}$$

One can readily verify that

$$\begin{aligned}
&\tfrac{7}{4} - P(+++|000) - \tfrac{1}{16}M - \tfrac{5}{16}M' \\
&\quad = \tfrac{1}{4}P(++-|000) + \tfrac{1}{4}P(+-+|000) + \tfrac{1}{4}P(-++|000) \\
&\qquad + \tfrac{3}{4}P(---|000)
\end{aligned}$$

$$
\begin{aligned}
&+ P(+-+|001) + P(-++|001) + \tfrac{1}{2}P(---|001) \\
&+ P(++-|010) + P(-++|010) + \tfrac{1}{2}P(---|010) \\
&+ P(++-|100) + P(+-+|100) + \tfrac{1}{2}P(---|100) \\
&+ \tfrac{1}{2}P(+--|011) + \tfrac{1}{2}P(-+-|101) \\
&+ \tfrac{1}{2}P(--+|110) \\
&+ \tfrac{1}{4}P(+++|111) + \tfrac{3}{4}P(+--|111) \\
&+ \tfrac{3}{4}P(-+-|111) + \tfrac{3}{4}P(--+|111) \\
&\geq 0
\end{aligned} \tag{C.35}
$$

and

$$
\begin{aligned}
&\tfrac{7}{4} - P(---|000) - \tfrac{1}{16}M - \tfrac{5}{16}M' \\
&\quad = \tfrac{1}{2}P(+++|000) \\
&\quad\quad + \tfrac{1}{4}P(++-|001) + P(+-+|001) + P(-++|001) \\
&\quad\quad + \tfrac{1}{4}P(---|001) \\
&\quad\quad + P(++-|010) + \tfrac{1}{4}P(+-+|010) + P(-++|010) \\
&\quad\quad + \tfrac{1}{4}P(---|010) \\
&\quad\quad + P(++-|100) + P(+-+|100) + \tfrac{1}{4}P(-++|100) \\
&\quad\quad + \tfrac{1}{4}P(---|100) \\
&\quad\quad + \tfrac{1}{4}P(+++|011) + \tfrac{1}{4}P(+--|011) \\
&\quad\quad + \tfrac{1}{4}P(+++|101) + \tfrac{1}{4}P(-+-|101) \\
&\quad\quad + \tfrac{1}{4}P(+++|110) + \tfrac{3}{4}P(--+|110) \\
&\quad\quad + \tfrac{1}{2}P(---|110) \\
&\quad\quad + P(+--|111) + P(-+-|111) + \tfrac{1}{2}P(--+|111) \\
&\quad\quad \geq 0
\end{aligned} \tag{C.36}
$$

under the no-signalling constraints.

The bounds given here are the tightest that can be derived given that there are no-signalling distributions for which:

$$
\big(G(A|0), M, M'\big) \in \big\{\big(1,\ 0,\ \pm'4\big), \big(1,\ \pm4,\ 0\big), \big(\tfrac{1}{2},\ \pm4,\ \pm'4\big)\big\}, \tag{C.37}
$$

$$
\begin{aligned}
\big(G(AB|00), M, M'\big) \in \big\{&\big(1,\ \pm2,\ \pm'2\big), \big(\tfrac{1}{2},\ \pm2,\ \pm'4\big), \\
&\big(\tfrac{1}{2},\ \pm4,\ \pm'2\big), \big(\tfrac{1}{4},\ \pm4,\ \pm'4\big)\big\},
\end{aligned} \tag{C.38}
$$

and

$$\big(G(ABC|111), M, M'\big) \in \big\{\big(1,\ \pm 2,\ \pm' 2\big), \big(\tfrac{1}{2},\ \pm 4,\ \pm' 2\big), \big(\tfrac{1}{2},\ 0,\ \pm' 4\big), \big(\tfrac{1}{4},\ \pm 4,\ \pm' 4\big)\big\}. \tag{C.39}$$

The only case that might not be immediately obvious is that there are no-signalling distributions for which simultaneously $G(AB|00) = 1/2$, $M = \pm 4$, and $M' = \pm' 2$; these can be attained with vertices of class 34 according to the classification used in table 1 of [PBS11].

C.2 Possible bound for $n > 3$ parties

In Section 5.3.3 we showed that the upper bound (5.16) on the two-party guessing probability $G(AB|00)$ is tight and the nonlinear part $M \geq 3$ can be attained if the parties measure σ_{x} and σ_{y} on a state of the form

$$|\Psi\rangle = \lambda\,|{+}{+}{+}\rangle + \mu\big(|{+}{-}{-}\rangle + |{-}{+}{-}\rangle + |{-}{-}{+}\rangle\big)\,. \tag{C.40}$$

We mention a possible extension here for the n-partite Mermin correlator

$$M_n = \mathrm{Re}\Big[\Big\langle \prod_{p=1}^{n} \big(A_0^{(p)} + iA_1^{(p)}\big)\Big\rangle\Big], \tag{C.41}$$

where $A_x^{(p)}$ are the pth party's measurement operators, whose local and quantum bounds are respectively [Mer90]

$$L_n = \begin{cases} 2^{(n-1)/2} & \text{if } n \text{ odd} \\ 2^{n/2} & \text{if } n \text{ even} \end{cases} \tag{C.42}$$

and

$$Q_n = 2^{n-1} \tag{C.43}$$

(although the local bound $M_n \leq L_n$ is a facet of the local polytope only for odd n).

The obvious generalisation of the strategy of section 5.3.3 is for the n parties to measure

$$A_0^{(p)} = \sigma_{\mathrm{x}}\,, \qquad\qquad A_2^{(1)} = \sigma_{\mathrm{y}} \tag{C.44}$$

on an n-partite state of the form

$$|\Psi\rangle = \lambda\,|+\rangle^{\otimes n} + \mu \sum_{\boldsymbol{s}\in\mathcal{S}} |\boldsymbol{s}\rangle\ , \tag{C.45}$$

where $\mathcal{S} \subset \{+,-\}^{\times n}$ is the subset of all vectors of n signs with a nonzero even number of minuses. The state is normalised if

$$\lambda^2 + (Q_n - 1)\mu^2 = 1\,. \tag{C.46}$$

In terms of λ and μ, the probability that the first $n-1$ parties (or all n of them, for that matter) obtain the result '+' if they measure σ_{x} is

$$P_{\boldsymbol{A}|\boldsymbol{X}}(+|\mathbf{1}) = \lambda^2 \tag{C.47}$$

and the Mermin expectation value is

$$M_n = \big(\lambda + (Q_n - 1)\mu\big)^2\,. \tag{C.48}$$

Relating $P_{\boldsymbol{A}|\boldsymbol{X}}(+|\mathbf{1}) = \lambda^2$ to M_n yields the dependence $P_{\boldsymbol{A}|\boldsymbol{X}}(\mathbf{1}|\mathbf{1}) = P_n(M_n)$, where

$$\begin{aligned} P_n(M_n) = 1 - \frac{1}{Q_n} - \frac{Q_n - 2}{{Q_n}^2} M_n \\ + 2\frac{\sqrt{Q_n - 1}}{{Q_n}^2}\sqrt{M_n(Q_n - M_n)}\,. \end{aligned} \tag{C.49}$$

By suitably mixing this strategy with a deterministic strategy with $P_{\boldsymbol{A}|\boldsymbol{X}}(\mathbf{1}|\mathbf{1}) = 1$ and $M_n = L_n$, we obtain a strategy for which the guessing probability and Mermin expectation value are related by

$$G(\boldsymbol{A}|\mathbf{0}) = \begin{cases} P_n(M_n) & \text{if } M_n \geq M_n^{\mathrm{th}} \\ \Gamma_n(M_n) & \text{if } M_n \leq M_n^{\mathrm{th}} \end{cases}\ , \tag{C.50}$$

where

$$\Gamma_n(M_n) = \frac{L_n(Q_n - 1) - (L_n - 1)M_n}{L_n(Q_n - L_n)}\,. \tag{C.51}$$

The threshold M_n^{th} in (C.50) is the point where the linear interpolation $\Gamma_n(M_n)$ coincides with the curve $P_n(M_n)$ and their derivatives are the same. This occurs at

$$M_n^{\mathrm{th}} = \frac{{L_n}^2(Q_n - 1)}{{L_n}^2 - 2L_n + Q_n}\,, \tag{C.52}$$

at which point

$$G(\boldsymbol{A}|\mathbf{0}) = \frac{Q_n - 1}{{L_n}^2 - 2L_n + Q_n}\,. \tag{C.53}$$

For odd n, we remark that $L_n = \sqrt{Q_n}$ and in that case (C.52) reduces to the average $M_n^{\text{th}} = (L_n + Q_n)/2$ of the local and quantum bounds.

The strategy we have described here shows that the upper bound on the guessing probability cannot be better than (C.50). For $n = 4$ and 5 parties, some numerical tests we carried out seemed to support that the upper bound on the guessing probability coincides with (C.50), although we did not attempt to prove this.

D. Appendix of Chapter 6

The tilted CHSH expression [AMP12] reads:

$$I_\beta = \beta\langle A\rangle + \langle AB\rangle + \langle AB'\rangle + \langle A'B\rangle - \langle A'B'\rangle, \tag{D.1}$$

where A, A', B, and B' are measurement operators with $-\mathbb{1}_\mathrm{A} \leq A, A' \leq \mathbb{1}_\mathrm{A}$ acting on $\mathcal{H}_\mathrm{A}$ and $-\mathbb{1}_\mathrm{B} \leq B, B' \leq \mathbb{1}_\mathrm{B}$ acting on $\mathcal{H}_\mathrm{B}$. For $0 \leq \beta < 2$, I_β satisfies the tight quantum bound

$$I_\beta \leq 2\sqrt{2}\sqrt{1+\beta^2/4} \tag{D.2}$$

which is strictly higher than the local bound $I_\beta \leq |\beta| + 2$. Eq. (D.2) can be attained with equality if (for example) Alice and Bob measure

$$A = \sigma_\mathrm{z}\,, \qquad\qquad A' = \sigma_\mathrm{x} \tag{D.3}$$

and

$$B = \cos\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_\mathrm{z} + \sin\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_\mathrm{x}\,, \qquad B' = \cos\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_\mathrm{z} - \sin\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_\mathrm{x} \tag{D.4}$$

on the two-qubit pure state

$$|\psi_\beta\rangle = \cos\!\left(\tfrac{\theta_\beta}{2}\right)|00\rangle + \sin\!\left(\tfrac{\theta_\beta}{2}\right)|11\rangle\,, \tag{D.5}$$

where μ_β and θ_β are related to β by

$$\sin(\theta_\beta) = \sqrt{\frac{1-\beta^2/4}{1+\beta^2/4}}\,, \qquad \cos(\theta_\beta) = \sqrt{\frac{2\,\beta^2/4}{1+\beta^2/4}}\,, \tag{D.6}$$

$$\sin\!\left(\tfrac{\mu_\beta}{2}\right) = \sqrt{\frac{1-\beta^2/4}{2}}\,, \qquad \cos\!\left(\tfrac{\mu_\beta}{2}\right) = \sqrt{\frac{1+\beta^2/4}{2}}\,. \tag{D.7}$$

Inversely, β and μ_β are related to θ_β by

$$\beta = \frac{2\cos(\theta_\beta)}{\sqrt{1+\sin(\theta_\beta)^2}}\,, \qquad \tan\left(\tfrac{\mu_\beta}{2}\right) = \sin(\theta_\beta)\,. \tag{D.8}$$

This tells us what value of β and what measurements to do on Bob's side if we're aiming to identify a state for some given angle θ_β.

The purpose of this Appendix is to establish the following self-testing result: if the quantum bound (D.2) is attained with equality then there is a choice of basis in which the state has the form:

$$\rho = \psi_\beta \otimes \sigma_{\text{junk}}\,, \tag{D.9}$$

where $\psi_\beta = |\psi_\beta\rangle\langle\psi_\beta|$ is the pure qubit state above, and Alice's measurements are:

$$A = \sigma_{\text{z}} \otimes \mathbb{1} \,\oplus\, A_\perp\,, \tag{D.10}$$

$$A' = \sigma_{\text{x}} \otimes \mathbb{1} \,\oplus\, A'_\perp\,, \tag{D.11}$$

where $\sigma_{\text{z}}\otimes\mathbb{1}$ and $\sigma_{\text{x}}\otimes\mathbb{1}$ act only on the support of the marginal state $\rho_{\text{A}} = \text{Tr}_{\text{B}}[\rho]$ on Alice's side and $A_\perp$ and $A'_\perp$ act only on its orthogonal complement in $\mathcal{H}_{\text{A}}$.

This result follows mostly from [AMP12]. We proceed by proving progressively more general self-testing results, first restricting to projective measurements on a bipartite pure qubit state, then generalising to arbitrary dimension using the Jordan lemma, then explicitly allowing for an underlying mixed state and non-projective measurements.

D.1 Qubit systems

The most general two-qubit pure state has the form:

$$|\psi\rangle = \cos\left(\tfrac{\theta}{2}\right)|00\rangle + \sin\left(\tfrac{\theta}{2}\right)|11\rangle\,, \tag{D.12}$$

for $0 \leq \theta \leq \pi/2$, in its Schmidt decomposition, while the most general projective measurements worth considering are:

$$A = \boldsymbol{a}\cdot\boldsymbol{\sigma}\,, \qquad B = \boldsymbol{b}\cdot\boldsymbol{\sigma}\,, \tag{D.13}$$

$$A' = \boldsymbol{a}'\cdot\boldsymbol{\sigma}\,, \qquad B' = \boldsymbol{b}'\cdot\boldsymbol{\sigma} \tag{D.14}$$

with $\|\boldsymbol{a}\| = \|\boldsymbol{a}'\| = \|\boldsymbol{b}\| = \|\boldsymbol{b}'\| = 1$, since we can't exceed the classical bound if any of the measurements are $\pm\mathbb{1}$, and where $\boldsymbol{\sigma} = (\sigma_{\mathrm{x}}, \sigma_{\mathrm{y}}, \sigma_{\mathrm{z}})$. We remind that the density operator associated with the state (D.12) can be written:

$$\psi = \frac{1}{4}\Big[\mathbb{1}\otimes\mathbb{1} + \cos(\theta)\big(\sigma_{\mathrm{z}}\otimes\mathbb{1} + \mathbb{1}\otimes\sigma_{\mathrm{z}}\big) + \sin(\theta)\big(\sigma_{\mathrm{x}}\otimes\sigma_{\mathrm{x}} - \sigma_{\mathrm{y}}\otimes\sigma_{\mathrm{y}}\big) + \sigma_{\mathrm{z}}\otimes\sigma_{\mathrm{z}}\Big]. \quad \text{(D.15)}$$

We write the expectation value of I_β as

$$I_\beta = \beta\cos(\theta)a_{\mathrm{z}} + S \quad \text{(D.16)}$$

where

$$S = \big\langle A(B + B') + A'(B - B')\big\rangle = \boldsymbol{a}\cdot\mathbf{T}(\boldsymbol{b} + \boldsymbol{b}') + \boldsymbol{a}'\cdot\mathbf{T}(\boldsymbol{b} - \boldsymbol{b}') \quad \text{(D.17)}$$

and

$$\mathbf{T} = \begin{bmatrix} \sin(\theta) & 0 & 0 \\ 0 & -\sin(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad \text{(D.18)}$$

Substituting now

$$\boldsymbol{b} + \boldsymbol{b}' = 2\cos\big(\tfrac{\mu}{2}\big)\boldsymbol{b}_+\,, \qquad \boldsymbol{b} - \boldsymbol{b}' = 2\sin\big(\tfrac{\mu}{2}\big)\boldsymbol{b}_-\,, \quad \text{(D.19)}$$

where $\boldsymbol{b}_\pm$ are normalised and orthogonal and we take $\cos\big(\tfrac{\mu}{2}\big), \sin\big(\tfrac{\mu}{2}\big) \geq 0$,

$$\begin{aligned} S &= 2\cos\big(\tfrac{\mu}{2}\big)\,\boldsymbol{a}\cdot\mathbf{T}\boldsymbol{b}_+ + 2\sin\big(\tfrac{\mu}{2}\big)\,\boldsymbol{a}'\cdot\mathbf{T}\boldsymbol{b}_- \\ &\leq 2\cos\big(\tfrac{\mu}{2}\big)\,\|\mathbf{T}\boldsymbol{b}_+\| + 2\sin\big(\tfrac{\mu}{2}\big)\,\|\mathbf{T}\boldsymbol{b}_-\| \\ &\leq 2\sqrt{\|\mathbf{T}\boldsymbol{b}_+\|^2 + \|\mathbf{T}\boldsymbol{b}_-\|^2} \\ &= 2\sqrt{\mathrm{Tr}\Big[\mathbf{T}^2\big(\boldsymbol{b}_+{\boldsymbol{b}_+}^T + \boldsymbol{b}_-{\boldsymbol{b}_-}^T\big)\Big]} \\ &\leq 2\sqrt{1 + \sin(\theta)^2}\,, \end{aligned} \quad \text{(D.20)}$$

where the last line follows from the inequality discussed in Appendix D.5. Using this in (D.16),

$$\begin{aligned} I_\beta &\leq \beta\cos(\theta)a_{\mathrm{z}} + 2\sqrt{1 + \sin(\theta)^2} \\ &\leq \beta\cos(\theta) + 2\sqrt{1 + \sin(\theta)^2} \\ &\leq 2\sqrt{2}\sqrt{1 + \beta^2/4}\,. \end{aligned} \quad \text{(D.21)}$$

In order to attain the quantum bound $I_\beta = 2\sqrt{2}\sqrt{1+\beta^2/4}$, all of the inequalities used to get from (D.16) to (D.21) must hold with equality. Working backwards, we extract that

$$2\cos(\theta) = \beta\sqrt{1+\sin(\theta)^2}\,, \tag{D.22}$$
$$\boldsymbol{a} = \mathbf{1}_{\mathrm{z}}\,, \tag{D.23}$$
$$\boldsymbol{b}_+ = \mathbf{1}_{\mathrm{z}}\,, \tag{D.24}$$
$$\boldsymbol{b}_- = \cos(\varphi)\mathbf{1}_{\mathrm{x}} - \sin(\varphi)\mathbf{1}_{\mathrm{y}}\,, \tag{D.25}$$
$$\boldsymbol{a}' = \cos(\varphi)\mathbf{1}_{\mathrm{x}} + \sin(\varphi)\mathbf{1}_{\mathrm{y}}\,, \tag{D.26}$$
$$\cos\left(\tfrac{\mu}{2}\right)\sin(\theta) = \sin\left(\tfrac{\mu}{2}\right). \tag{D.27}$$

Under the convention $\beta > 0$ and $0 \leq \theta_\beta, \frac{\mu_\beta}{2} \leq \frac{\pi}{2}$ that we are working with, these imply the relations (D.6) and (D.7) for θ_β and μ_β given above. The remaining undetermined parameter φ can be set to 0 e.g. with the phase changes $|1\rangle_{\mathrm{A}} \mapsto e^{i\varphi}|1\rangle_{\mathrm{A}}$ and $|1\rangle_{\mathrm{B}} \mapsto e^{-i\varphi}|1\rangle_{\mathrm{B}}$, under which the Schmidt decomposition is invariant.

In the derivation above we started by expressing $|\psi\rangle$ in its Schmidt decomposition and have shown that, if the quantum bound is attained, the measurements must satisfy:

$$A_0 = \sigma_{\mathrm{z}}\,, \qquad A_1 = \cos(\varphi)\sigma_{\mathrm{x}} + \sin(\varphi)\sigma_{\mathrm{y}}, \tag{D.28}$$

and

$$B_0 + B_1 \propto \sigma_{\mathrm{z}}, \qquad B_0 - B_1 \propto \cos(\varphi)\sigma_{\mathrm{x}} - \sin(\varphi)\sigma_{\mathrm{y}} \tag{D.29}$$

with respect to the Schmidt basis. It is important to note that the converse also holds: if the quantum bound is attained with measurements satisfying these conditions then the state must be exactly $|\psi_\beta\rangle = \cos\left(\frac{\theta_\beta}{2}\right)|00\rangle + \sin\left(\frac{\theta_\beta}{2}\right)|11\rangle$. The reasoning is thus the following: if the quantum bound is attained with qubits, then there is a choice of bases in which $A_0 = \sigma_{\mathrm{z}}$, $A_1 = \sigma_{\mathrm{x}}$, $B_0 + B_1 \propto \sigma_{\mathrm{z}}$, and $B_0 - B_1 \propto \sigma_{\mathrm{x}}$, and this then determines that the state is $|\psi_\beta\rangle$ with respect to that choice of the bases.

D.2 Arbitrary dimension

According to the Jordan lemma the measurement operators A, A' and B, B' can be block diagonalised in their respective Hilbert spaces into blocks no larger

than 2×2. We express the block diagonalisation as

$$A = \sum_j A_j \otimes |j\rangle\langle j| \oplus A_\perp \,, \tag{D.30}$$

$$A' = \sum_j A'_j \otimes |j\rangle\langle j| \oplus A'_\perp \,, \tag{D.31}$$

$$B = \sum_k B_k \otimes |k\rangle\langle k| \oplus B_\perp \,, \tag{D.32}$$

$$B' = \sum_k B'_k \otimes |k\rangle\langle k| \oplus B'_\perp \,, \tag{D.33}$$

where A_j, A'_j, B_k, and B'_k are 2×2 operators and the operators with '$\perp$' as subscript collectively denote any 1×1 blocks. Note that this implies $[A_\perp, A'_\perp] = 0$ and $[B_\perp, B'_\perp] = 0$. With respect to this splitting of the Hilbert space, we can express an arbitrary pure state as

$$|\Psi\rangle = \bigoplus_{mn} \sqrt{p_{mn}}\,|\psi_{mn}\rangle \tag{D.34}$$

where the indices $m, n \in \{2, \perp\}$ indicate whether the state is in the subspace containing the 2×2 or 1×1 blocks in $\mathcal{H}_\mathrm{A}$ and $\mathcal{H}_\mathrm{B}$. The expectation value of I_β splits accordingly as

$$I_\beta = \sum_{mn} p_{mn} I_\beta^{(mn)} \,. \tag{D.35}$$

In order to attain the quantum bound $I_\beta = 2\sqrt{2}\sqrt{1+\beta^2/4}$ we must have $I_\beta^{(mn)} = 2\sqrt{2}\sqrt{1+\beta^2/4}$ for each m, n for which $p_{mn} \neq 0$. However, except for $(m, n) = (2, 2)$, $I_\beta^{(mn)}$ is limited to the classical bound since the measurements on Alice's and/or Bob's side commute in the corresponding subspace. Thus, all of the support of $|\Psi\rangle$ must be in the subspace of $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$ containing the 2×2 blocks on both sides.

With respect to the 2×2 blocks, the state can be expressed as

$$|\Psi\rangle = \sum_{jk} \sqrt{q_{jk}}\,|\phi_{jk}\rangle\,|j\rangle\,|k\rangle \tag{D.36}$$

and the value of I_β, accordingly,

$$\begin{aligned} I_\beta &= \sum_{jk} q_{jk} \langle\phi_{jk}| \left(A_j(B_k + B'_k) + A'_j(B_k - B'_k)\right) |\phi_{jk}\rangle \,, \\ &= \sum_{jk} q_{jk} I_\beta^{(jk)} \,. \end{aligned} \tag{D.37}$$

Again, in order to attain the quantum bound, for each contribution (j,k), we must have either $I_\beta^{(jk)} = 2\sqrt{2}\sqrt{1+\beta^2/4}$ or $q_{jk} = 0$. We can first get rid of the parts of Alice's and Bob's Hilbert spaces that don't contain $|\Psi\rangle$: if there are any j's such that $\forall k,\ q_{jk} = 0$ or any k's such that $\forall j,\ q_{jk} = 0$, we absorb the corresponding blocks $A_j \otimes |j\rangle\langle j|$ and $A'_j \otimes |j\rangle\langle j|$ or $B_k \otimes |k\rangle\langle k|$ and $B'_k \otimes |k\rangle\langle k|$ respectively into $A_\perp$ and $A'_\perp$ or $B_\perp$ and $B'_\perp$. For the remaining blocks, for each j there is at least one k and for each k at least one j such that $q_{jk} \neq 0$ and we must have $I_\beta^{(jk)} = 2\sqrt{2}\sqrt{1+\beta^2/4}$. Following the remark at the end of the last Section, there is a choice of bases in which, for all the remaining j and k,

$$A_j = \sigma_{\mathrm{z}}\,, \qquad\qquad A'_j = \sigma_{\mathrm{x}}\,, \tag{D.38}$$

and

$$B_k + B'_k = 2\cos\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_{\mathrm{z}}, \qquad\qquad B_k - B'_k = 2\sin\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_{\mathrm{x}}. \tag{D.39}$$

This in turn implies $|\psi_{jk}\rangle = |\psi_\beta\rangle$ for all the remaining j,k for which $q_{jk} \neq 0$. We can also choose to set $|\psi_{jk}\rangle$ for the others since if $q_{jk} = 0$ then $\sqrt{q_{jk}}\,|\psi_{jk}\rangle = 0$ regardless of what $|\psi_{jk}\rangle$ is. We thus obtain that the state and measurements, in a suitable choice of the bases, are

$$|\Psi\rangle = |\psi_\beta\rangle \otimes |\mathrm{junk}\rangle\ , \tag{D.40}$$

with $|\mathrm{junk}\rangle = \sum_{jk} \sqrt{q_{jk}}\,|j\rangle\,|k\rangle$, and

$$A = \sigma_{\mathrm{z}} \otimes \mathbb{1}\ \oplus\ A_\perp\,, \tag{D.41}$$

$$A' = \sigma_{\mathrm{x}} \otimes \mathbb{1}\ \oplus\ A'_\perp\,, \tag{D.42}$$

$$B = \left(\cos\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_{\mathrm{z}} + \sin\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_{\mathrm{x}}\right) \otimes \mathbb{1}\ \oplus\ B_\perp\,, \tag{D.43}$$

$$B' = \left(\cos\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_{\mathrm{z}} - \sin\!\left(\tfrac{\mu_\beta}{2}\right)\sigma_{\mathrm{x}}\right) \otimes \mathbb{1}\ \oplus\ B'_\perp\,, \tag{D.44}$$

where only the first terms $(\ldots) \otimes \mathbb{1}$ act on the parts of $\mathcal{H}_{\mathrm{A}}$ and $\mathcal{H}_{\mathrm{B}}$ containing $|\Psi\rangle$.

D.3 Mixed states

The derivation up to this point easily adapts to allow for mixed states, since an arbitrary mixed state can be expressed as a convex sum

$$\rho = \sum_s p_s \Psi_s \tag{D.45}$$

of pure states. In order to attain the quantum bound for I_β, it must be attained with each pure state $|\Psi_s\rangle$. Following the reasoning of the previous Section, we deduce that all the $|\Psi_s\rangle$ are in the subspace of $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$ containing the 2×2 measurement operator blocks and have the form

$$|\Psi_s\rangle = \sum_{jk} \sqrt{q^s_{jk}}\, |\phi^s_{jk}\rangle\, |j\rangle\, |k\rangle \ . \tag{D.46}$$

The only difference is that we only discard the blocks j for which $\forall k, s,\ q^s_{jk} = 0$ and k for which $\forall j, s,\ q^s_{jk} = 0$. We then obtain

$$|\Psi_s\rangle = |\psi_\beta\rangle \otimes |\mathrm{junk}_s\rangle \tag{D.47}$$

and, in turn,

$$\rho = \psi_\beta \otimes \sigma_\mathrm{junk}\,, \tag{D.48}$$

where σ_junk is a (not necessarily pure) state

$$\sigma_\mathrm{junk} = \sum_s p_s\ |\mathrm{junk}_s\rangle\, \langle \mathrm{junk}_s|\ . \tag{D.49}$$

D.4 General measurements

In general, measurements with only two outcomes can be expressed as convex sums of projective measurements. For the measurement operators we may write

$$A = \sum_j p_j A_j\,, \qquad A' = \sum_j p_j A'_j\,, \tag{D.50}$$

$$B = \sum_k q_k B_k\,, \qquad B' = \sum_k q_k B_k\,, \tag{D.51}$$

with ${A_j}^2 = {A'_j}^2 = \mathbb{1}_\mathrm{A}$ and ${B_k}^2 = {B'_k}^2 = \mathbb{1}_\mathrm{B}$. I_β then decomposes as

$$I_\beta = \sum_{jk} p_j q_k\, I_\beta^{(jk)} \tag{D.52}$$

with

$$I_\beta^{(jk)} = \left\langle \beta A_j + A_j B_k + A_j B'_k + A'_j B_k - A'_j B'_k \right\rangle . \tag{D.53}$$

Obviously, if the quantum bound is attained then all the $I_\beta^{(jk)}$s have to attain it individually. In particular, for $i = j = 1$, the results of the previous Sections imply that there is a choice of the bases in which the underlying state is

$$\rho = \psi_\beta \otimes \sigma_{\text{junk}} \tag{D.54}$$

and

$$A_1 = \sigma_{\text{z}} \otimes \mathbb{1} \ \oplus \ A_\perp^{(1)} , \tag{D.55}$$

$$A_1' = \sigma_{\text{x}} \otimes \mathbb{1} \ \oplus \ A'^{(1)}_\perp , \tag{D.56}$$

$$B_1 = \Big(\cos\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\text{z}} + \sin\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\text{x}}\Big) \otimes \mathbb{1} \ \oplus \ B_\perp^{(1)} , \tag{D.57}$$

$$B_1' = \Big(\cos\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\text{z}} - \sin\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\text{x}}\Big) \otimes \mathbb{1} \ \oplus \ B'^{(1)}_\perp . \tag{D.58}$$

Consider now $I_\beta^{(j1)}$ for $j \neq 1$. We can write it as

$$\begin{aligned} I_\beta^{(j1)} &= \big\langle A_j\big(\beta\mathbb{1} + B_1 + B_1'\big)\big\rangle + \big\langle A_j'\big(B_1 - B_1'\big)\big\rangle \\ &= \text{Tr}\big[A_j\,(\tilde\rho_+ \otimes \sigma_{\text{A}})\big] + \text{Tr}\big[A_j'\,(\tilde\rho_- \otimes \sigma_{\text{A}})\big] \end{aligned} \tag{D.59}$$

where

$$\begin{aligned} \tilde\rho_+ = &\Big(\tfrac{\beta}{2} + \cos\big(\tfrac{\mu_\beta}{2}\big)\cos(\theta_\beta)\Big)\mathbb{1} \\ &+ \Big(\tfrac{\beta}{2}\cos(\theta_\beta) + \cos\big(\tfrac{\mu_\beta}{2}\big)\Big)\sigma_{\text{z}} , \end{aligned} \tag{D.60}$$

$$\tilde\rho_- = \sin\big(\tfrac{\mu_\beta}{2}\big)\sin(\theta_\beta)\sigma_{\text{x}} , \tag{D.61}$$

and σ_{A} is the marginal of σ_{junk} on Alice's side. Using the relations (D.6) and (D.7) for θ_β and μ_β in terms of β,

$$\tilde\rho_+ \otimes \sigma_{\text{A}} = \frac{1}{2}\left[2\beta\,\mathbb{1} + \frac{\sqrt{2}\big(1 + 3\,\beta^2/4\big)}{\sqrt{1+\beta^2/4}}\,\sigma_{\text{z}}\right] \otimes \sigma_{\text{A}} , \tag{D.62}$$

$$\tilde\rho_- \otimes \sigma_{\text{A}} = \frac{1}{2}\,\frac{\sqrt{2}\big(1 - \beta^2/4\big)}{\sqrt{1+\beta^2/4}}\,\sigma_{\text{x}} \otimes \sigma_{\text{A}} . \tag{D.63}$$

In order for the traces in (D.59) to reach their maximal values, A_j and A_j' must be diagonal in the same bases as the operators $\tilde\rho_+ \otimes \sigma_{\text{A}}$ and $\tilde\rho_- \otimes \sigma_{\text{A}}$ that they are multiplied with. Note that $\tilde\rho_+$ in (D.62) has a negative eigenvalue for $\beta < 2$; a little algebra shows that

$$\frac{\sqrt{2}\big(1 + 3\,\beta^2/4\big)}{\sqrt{1+\beta^2}} > 2\beta \tag{D.64}$$

rearranges to and is implied by $(1-\beta^2/4)^2 > 0$. We can thus infer that

$$A_j = \sigma_{\rm z} \otimes \mathbb{1} \ \oplus \ A_\perp^{(j)} \,, \qquad\qquad A'_j = \sigma_{\rm x} \otimes \mathbb{1} \ \oplus \ A'^{(j)}_\perp \tag{D.65}$$

for all j, and that A and A' have the form

$$A = \sigma_{\rm z} \otimes \mathbb{1} \ \oplus \ A_\perp \,, \qquad\qquad A' = \sigma_{\rm x} \otimes \mathbb{1} \ \oplus \ A'_\perp \tag{D.66}$$

where $A_\perp = \sum_j p_j A_\perp^{(j)}$ and $A'_\perp = \sum_j p_j A'^{(j)}_\perp$ are bounded between $-\mathbb{1}$ and $\mathbb{1}$.

Applying the same approach to $I_\beta^{(1k)}$ for $k \neq 1$ we can similarly deduce that

$$B = \Big(\cos\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\rm z} + \sin\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\rm x}\Big) \otimes \mathbb{1} \ \oplus \ B_\perp \,, \tag{D.67}$$

$$B' = \Big(\cos\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\rm z} - \sin\big(\tfrac{\mu_\beta}{2}\big)\sigma_{\rm x}\Big) \otimes \mathbb{1} \ \oplus \ B'_\perp \tag{D.68}$$

with $B_\perp = \sum_k q_k B_\perp^{(k)}$ and $B'_\perp = \sum_k q_k B'^{(k)}_\perp$.

D.5 Eigenvalue von Neumann trace inequality

If A and B are Hermitian operators then:

1. The trace of their product respects

$$\mathrm{Tr}[AB] \leq \sum_k a_k b_k \tag{D.69}$$

   where a_k and b_k are the eigenvalues of A and B ordered from largest to smallest.

2. (D.69) is attained with equality if and only if there is a basis in which A and B are both diagonal and the ordering of their eigenvalues by magnitude match.

Eq. (D.69) is just a version of von Neumann's trace inequality for Hermitian operators. We go over the proof here just to explicitly confirm point 2.

We write

$$A = \sum_k a_k \left|\alpha_k\right\rangle \left\langle\alpha_k\right| \,, \qquad\qquad B = \sum_k b_k \left|\beta_k\right\rangle \left\langle\beta_k\right| \,, \tag{D.70}$$

where we choose the labelling such that a_k and b_k are ordered from largest to smallest. Then

$$\mathrm{Tr}[AB] = \sum_{kk'} a_k b_{k'} \left|\langle \alpha_k \rangle \, \beta_{k'}\right|^2 . \tag{D.71}$$

Here, $\left|\langle \alpha_k \rangle \, \beta_{k'}\right|^2$ are the elements of a doubly stochastic matrix which, according to the Birkhoff-von Neumann theorem, can be expressed as the convex sum of permutation matrices, i.e., we can write

$$\left|\langle \alpha_k \rangle \, \beta_{k'}\right|^2 = \sum_{\pi} p_\pi \, \delta_{\pi(k),k'} \tag{D.72}$$

where the sum is taken over all permutations π of the set of indices $\{k\}$. So,

$$\mathrm{Tr}[AB] = \sum_{\pi} p_\pi \sum_k a_k b_{\pi(k)} \leq \sum_k a_k b_k \, .$$

In order for the upper bound to be attained with equality, we must have either $p_\pi = 0$ or $\sum_k a_k b_{\pi(k)} = \sum_k a_k b_k$ for every permutation π in the sum. The latter can only happen for permutations other than the identity permutation if some of the eigenvalues a_k or b_k are degenerate, in which case we can change the basis $\{|\alpha_k\rangle\}$ or $\{|\beta_k\rangle\}$ until we are only left with $|\langle \alpha_k \rangle \, \beta_{k'}|^2 = \delta_{kk'}$. For example, suppose $b_k = b_{k'}$ for some $k \neq k'$. Then we can force $\langle \alpha_k \rangle \, \beta_{k'} = 0$ by replacing

$$\begin{aligned} |\beta_k\rangle &\mapsto c^* \, |\beta_k\rangle + c'^* \, |\beta_{k'}\rangle \\ |\beta_{k'}\rangle &\mapsto -c' \, |\beta_k\rangle + c \, |\beta_{k'}\rangle \, , \end{aligned} \tag{D.74}$$

where

$$c = \frac{\langle \alpha_k \rangle \, \beta_k}{\sqrt{\left|\langle \alpha_k \rangle \, \beta_k\right|^2 + \left|\langle \alpha_k \rangle \, \beta_{k'}\right|^2}} \, , \tag{D.75}$$

$$c' = \frac{\langle \alpha_k \rangle \, \beta_{k'}}{\sqrt{\left|\langle \alpha_k \rangle \, \beta_k\right|^2 + \left|\langle \alpha_k \rangle \, \beta_{k'}\right|^2}} \, . \tag{D.76}$$

Bibliography

[ABG+07] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98(23):230501, 2007.

[ACP+16] A. Acín, D. Cavalcanti, E. Passaro, S. Pironio, and P. Skrzypczyk. Necessary detection efficiencies for secure quantum key distribution and bound randomness. *Phys. Rev. A*, 93:012319, Jan 2016.

[AFDF+18] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nat. Commun.*, 9(1):459, 2018.

[AFHTS12] R. Arnon-Friedman, E. Hänggi, and A. Ta-Shma. Towards the impossibility of non-signalling privacy amplification from time-like ordering constraints. arXiv:1205.3736, 2012.

[AFTS12] R. Arnon-Friedman and A. Ta-Shma. Limits of privacy amplification against nonsignaling memory attacks. *Phys. Rev. A*, 86(6):062333, 2012.

[AGCA12] L. Aolita, R. Gallego, A. Cabello, and A. Acín. Fully nonlocal, monogamous, and random genuinely multipartite quantum correlations. *Phys. Rev. Lett.*, 108:100401, 2012.

[AGR82] A. Aspect, P. Grangier, and G. Roger. Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities. *Phys. Rev. Lett.*, 49:91–94, 1982.

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

[APVW16] A. Acín, S. Pironio, T. Vértesi, and P. Wittek. Optimal randomness certification from one entangled bit. *Phys. Rev. A*, 93:040102, 2016.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore*, 1984.

[BBC95] C. H. Bennett, G. Brassard, and C. Cépreau. Generalized privacy amplification. *IEEE T. Inform. Theory*, 41(6):1915–1923, 1995.

[BCP+14] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86(2):419–478, 2014.

[Bel64] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(3):195–200, 1964.

[BGLP11] J.-D. Bancal, N. Gisin, Y.-C. Liang, and S. Pironio. Device-independent witnesses of genuine multipartite entanglement. *Phys. Rev. Lett.*, 106:250404, 2011.

[BHK05] J. Barrett, L. Hardy, and A. Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95(1):010503, 2005.

[BKG+18] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, M. J. Stevens, and L. K. Shalm. Experimentally generated randomness certified by the impossibility of superluminal signals. *Nature*, 556(7700):223–226, 2018.

[BKP06] J. Barrett, A. Kent, and S. Pironio. Maximally nonlocal and monogamous quantum correlations. *Phys. Rev. Lett.*, 97:170409, Oct 2006.

[BMP18] C. Bamps, S. Massar, and S. Pironio. Device-independent randomness generation with sublinear shared quantum resources. *Quantum*, 2:86, 2018.

[BP15] C. Bamps and S. Pironio. Sum-of-squares decompositions for a family of Clauser-Horne-Shimony-Holt-like inequalities and their application to self-testing. *Phys. Rev. A*, 91(5):052111, 2015.

[BPA18a] B. Bourdoncle, S. Pironio, and A. Acín. See ancillary files for arXiv:1807.04674, 2018.

[BPA18b] B. Bourdoncle, S. Pironio, and A. Acín. Quantifying the randomness of copies of noisy Popescu-Rohrlich correlations. *Phys. Rev. A*, 2018.

[BSS14] J.-D. Bancal, L. Sheridan, and V. Scarani. More randomness from the same data. *New J. Phys.*, 16(3):033011, 2014.

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

[CJ06] T. M. Cover and T. A. Joy. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[Col06] R. Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, University of Cambridge, 2006.

[CR12] R. Colbeck and R. Renner. Free randomness can be amplified. *Nat. Phys.*, 8:450, 2012.

[DFR16] F. Dupuis, O. Fawzi, and R. Renner. Entropy accumulation. arXiv:1607.01796, 2016.

[dlTHD+15] G. de la Torre, M. J. Hoban, C. Dhara, G. Prettico, and A. Acín. Maximally nonlocal theories cannot be maximally random. *Phys. Rev. Lett.*, 114:160502, 2015.

[DPA13] C. Dhara, G. Prettico, and A. Acín. Maximal quantum randomness in Bell tests. *Phys. Rev. A*, 88(5):052116, 2013.

[DPVR12] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan's extractor in the presence of quantum side information. *SIAM J. Comput.*, 41(4):915–940, 2012.

[Eke91] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67(6):661–663, 1991.

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

[FGS13] S. Fehr, R. Gelles, and C. Schaffner. Security and composability of randomness expansion from bell inequalities. *Phys. Rev. A*, 87:012335, 2013.

[FHSW10] M. Fitzi, E. Hänggi, V. Scarani, and S. Wolf. The non-locality of n noisy popescu–rohrlich boxes. *J. Phys. A - Math. Theor.*, 43(46):465305, 2010.

[Fin82] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.*, 48:291–295, 1982.

[GHSZ90] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger. Bell's theorem without inequalities. *Am. J. Phys.*, 58(12):1131–1143, 1990.

[GKW+18] K. T. Goh, J. Kaniewski, E. Wolfe, T. Vértesi, X. Wu, Y. Cai, Y.-C. Liang, and V. Scarani. Geometry of the quantum set of correlations. *Phys. Rev. A*, 97(2):022104, Feb 2018.

[GMdlT+13] R. Gallego, L. Masanes, G. de la Torre, C. Dhara, L. Aolita, and A. Acín. Full randomness from arbitrarily deterministic events. *Nat. Commun.*, 4:2654, 2013.

[GZ17] S. Gogioso and W. Zeng. Generalised Mermin-type non-locality arguments. arXiv:1702.01772, 2017.

[HBcvB99] M. Hillery, V. Bužek, and A. Berthiaume. Quantum secret sharing. *Phys. Rev. A*, 59:1829–1834, 1999.

[HBD+15] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682–686, 2015.

[HRW10] E. Hänggi, R. Renner, and S. Wolf. Efficient device-independent quantum key distribution. In *EUROCRYPT 2010: Advances in Cryptology*, 2010.

[HRW13] E. Hänggi, R. Renner, and S. Wolf. The impossibility of non-signaling privacy amplification. *Theor. Comput. Sci.*, 486:27 – 42, 2013.

[ILL89] R. Impagliazzo, L. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 12–24, 1989.

[KKI99] A. Karlsson, M. Koashi, and N. Imoto. Quantum entanglement for secret sharing and secret splitting. *Phys. Rev. A*, 59:162–168, 1999.

[KL51] S. Kullback and R. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

[KR11] R. König and R. Renner. Sampling of min-entropy relative to quantum knowledge. *IEEE T. Inform. Theory*, 57(7):4760–4787, 2011.

[KRS09] R. König, R. Renner, and C. Schaffner. The operational meaning of min- and max-entropy. *IEEE T. Inform. Theory*, 55(9):4337–4347, 2009.

[KW16] J. Kaniewski and S. Wehner. Device-independent two-party cryptography secure against sequential attacks. *New J. Phys.*, 18(5):055004, 2016.

[KZB17] E. Knill, Y. Zhang, and P. Bierhorst. Quantum randomness generation by probability estimation with classical side information. arXiv:1709.06159, 2017.

[KZF18] E. Knill, Y. Zhang, and H. Fu. Quantum probability estimation for randomness with quantum side information. arXiv:1806.04553, 2018.

[LHBR10] Y.-C. Liang, N. Harrigan, S. D. Bartlett, and T. Rudolph. Nonclassical correlations from randomly chosen local measurements. *Phys. Rev. Lett.*, 104:050401, Feb 2010.

[LRZ+18] P.-S. Lin, D. Rosset, Y. Zhang, J.-D. Bancal, and Y.-C. Liang. Device-independent point estimation from finite data and its application to device-independent property estimation. *Phys. Rev. A*, 97:032309, Mar 2018.

[LYL+18] Y. Liu, X. Yuan, M.-H. Li, W. Zhang, Q. Zhao, J. Zhong, Y. Cao, Y.-H. Li, L.-K. Chen, H. Li, T. Peng, Y.-A. Chen, C.-Z. Peng, S.-C.

Shi, Z. Wang, L. You, X. Ma, J. Fan, Q. Zhang, and J.-W. Pan. High-speed device-independent quantum random number generation without a detection loophole. *Phys. Rev. Lett.*, 120:010503, 2018.

[MAS06] V. Makarov, A. Anisimov, and J. Skaar. Effects of detector efficiency mismatch on security of quantum cryptosystems. *Phys. Rev. A*, 74(2):022313, 2006.

[MBL+13] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne. Device-independent entanglement quantification and related applications. *Phys. Rev. Lett.*, 111:030501, Jul 2013.

[Mer90] N. D. Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.*, 65:1838–1840, 1990.

[MP13] P. Mironowicz and M. Pawłowski. Robustness of quantum-randomness expansion protocols in the presence of noise. *Phys. Rev. A*, 88:032319, 2013.

[MPA11] L. Masanes, S. Pironio, and A. Acín. Secure device-independent quantum key distribution with causally independent measurement devices. *Nat. Commun.*, 2:238, 2011.

[MRC+14] L. Masanes, R. Renner, M. Christandl, A. Winter, and J. Barrett. Full security of quantum key distribution from no-signaling constraints. *IEEE T. Inform. Theory*, 60(8):4973–4986, 2014.

[MSP+17] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, v. Roučka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz. SymPy: symbolic computing in Python. *PeerJ Computer Science*, 3:e103, 2017.

[MY98] D. Mayers and A. Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 503, 1998.

[MY04] D. Mayers and A. Yao. Self testing quantum apparatus. *Quantum Info. Comput.*, 4(4):273–286, 2004.

[Nak10] M. Nakata. A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: SDPA-GMP, -QD and -DD. In *IEEE International Symposium on Computer-Aided Control System Design*, pages 29–34. IEEE, 2010.

[NPA07] M. Navascués, S. Pironio, and A. Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98(1):010401, 2007.

[NPA08] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New J. Phys.*, 10(7):073013, 2008.

[NSBSP18] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio. Device-independent randomness generation from several Bell estimators. *New J. Phys.*, 2018. arXiv:1611.00352v2.

[NSPS14] O. Nieto-Silleras, S. Pironio, and J. Silman. Using complete measurement statistics for optimal device-independent randomness evaluation. *New J. Phys.*, 16(1):013035, 2014.

[PAM+10] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, A. Manning, and C. Monroe. Random numbers certified by Bell's theorem. *Nature*, 464:1021–1024, 2010.

[PBS11] S. Pironio, J.-D. Bancal, and V. Scarani. Extremal correlations of the tripartite no-signaling polytope. *J. Phys. A - Math. Theor.*, 44(6):065303, 2011.

[PM13] S. Pironio and S. Massar. Security of practical private randomness generation. *Phys. Rev. A*, 87:012336, 2013.

[PMLA13] S. Pironio, L. Masanes, A. Leverrier, and A. Acín. Security of device-independent quantum key distribution in the bounded-quantum-storage model. *Phys. Rev. X*, 3:031007, Aug 2013.

[PR94] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24(3):379–385, 1994.

[PSV16] S. Pironio, V. Scarani, and T. Vidick. Focus on device independent quantum information. *New J. Phys.*, 18(10):100202, 2016.

[RBaH$^+$16] R. Ramanathan, F. G. S. L. Brandão, K. Horodecki, M. Horodecki, P. Horodecki, and H. Wojewódka. Randomness amplification under minimal fundamental assumptions on the devices. *Phys. Rev. Lett.*, 117:230501, 2016.

[Ren05] R. Renner. *Security of quantum key distribution.* PhD thesis, ETH Zurich, 2005.

[RGK05] R. Renner, N. Gisin, and B. Kraus. Information-theoretic security proof for quantum-key-distribution protocols. *Phys. Rev. A*, 72:012332, 2005.

[RM17] R. Ramanathan and P. Mironowicz. Trade-offs in multi-party Bell inequality violations in qubit networks. arXiv:1704.03790, 2017.

[RRMG17] M.-O. Renou, D. Rosset, A. Martin, and N. Gisin. On the inequivalence of the CH and CHSH inequalities due to finite statistics. *J. Phys. A - Math. Theor.*, 50(25):255301, 2017.

[RSA78] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126, 1978.

[RW04] R. Renner and S. Wolf. Smooth Rényi entropy and applications. In *IEEE Int. Symp. Info.*, page 233, 2004.

[ŠASA16] I. Šupić, R. Augusiak, A. S. Salavrakos, and A. Acín. Self-testing protocols based on the chained Bell inequalities. *New Journal of Physics*, 18(3):035013, 2016.

[SDP11] SDPA official page. `http://sdpa.sourceforge.net/`, 2011.

[SG01] V. Scarani and N. Gisin. Spectral decomposition of Bell's operators for qubits. *Journal of Physics A General Physics*, 34(30):6043, 2001.

[SPM13] J. Silman, S. Pironio, and S. Massar. Device-independent randomness generation in the presence of weak cross-talk. *Phys. Rev. Lett.*, 110:100504, 2013.

[SV84] M. Santha and U. V. Vazirani. Generating quasi-random sequences from slightly-random sources. In *25th Annual Symposium on Fondations of Computer Science*, pages 434–440, 1984.

[Sve87] G. Svetlichny. Distinguishing three-body from two-body nonseparability by a bell-type inequality. *Phys. Rev. D*, 35:3066–3069, 1987.

[SW16] B. Salwey and S. Wolf. Stronger attacks on causality-based key agreement. In *IEEE Int. Symp. Info.*, 2016. arXiv:1601.07833.

[Tre01] L. Trevisan. Extractors and peudo-random generators. *J. ACM*, 48(4), 2001.

[Tsi80] B. S. Tsirelson. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4(2):93–100, 1980.

[TSSR11] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner. Leftover hashing against quantum side information. *IEEE T. Inform. Theory*, 57(8):5524–5535, 2011.

[Val02] A. Valentini. Signal-locality in hidden-variables theories. *Phys. Lett. A*, 297(5–6):273 – 278, 2002.

[VV12] U. Vazirani and T. Vidick. Certifiable quantum dice: Or, true random number generation secure against quantum adversaries. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 61–76, New York, NY, USA, 2012. ACM.

[WBA18a] E. Woodhead, B. Bourdoncle, and A. Acín. Randomness versus nonlocality in the Mermin-Bell experiment with three parties. *Quantum*, 2:82, 2018.

[WBA18b] E. Woodhead, B. Bourdoncle, and A. Acín. See ancillary files for arxiv:1804.09733, 2018.

[WKB+19] E. Woodhead, J. Kaniewski, B. Bourdoncle, A. Salavrakos, J. Bowles, R. Augusiak, and A. Acín. Maximal randomness from partially entangled states. In preparation, 2019.

[Woo18] E. Woodhead. GitHub - ewoodhead/npa-hierarchy. `https://github.com/ewoodhead/npa-hierarchy`, 2018.

[WW01] R. F. Werner and M. M. Wolf. All-multipartite Bell-correlation inequalities for two dichotomic observables per site. *Phys. Rev. A*, 64:032112, 2001.