ICFO – The Institute of Photonic Sciences
Castelldefels (Barcelona), Spain

PhD thesis

From quantum foundations to quantum information protocols and back

Gonzalo de la Torre Carazo

July 9, 2015

Thesis supervisor Prof. Dr. Antonio Acín

To my family

Acknowledgements

The present PhD thesis is the result of more than four years of work at The Institute of Photonic Sciences. Many people have contributed to this thesis in one way or another and I want to use this opportunity to thank them sincerely as this thesis belongs to them as well.

First to Toni, for being such a great advisor. For sharing his scientific knowledge and passion as well as for his guidance, trust and patience during all these years. Thanks also to Lluis for greatly influencing the way I understood physics and being an excellent mentor during my initial steps as a researcher.

I would like to thank as well some of my closest collaborators: To Chirag for the great work and life ramblings which I miss. To Matty for being a very llanito friend that happens to be a great physicist too. To Ariel for walking in front of me in computability road and being a true Santa when needed. To Miguel for welcoming me in Bristol and sharing his boundless creativity with me. To Thinh for his friendliness and excellent work and to Jean Daniel for his friendship, patience and for sharing his 'pedagogical method' to science.

I am also indebted to other collaborators for the stimulating discussions and all that I learned from them: Markus Muller, Tony Short, Stefano Pironio, Giuseppe, Leandro, Gabriel and Santiago.

Special thanks to Valerio Scarani and the rest of the group for warmly welcoming me in their group and making me don't want to leave Singapore.

Thanks also to the Quantum Information Theory group at ICFO for the great atmosphere to talk about physics and have fun, for the darts, the futbolin and the laughs and their friendship. Big thanks go to Belencita (tucu tucu!), Rodrigo, Lars, Marti, Giuseppe, Leonardo, Flavio, Planeta, Augusto, Mafalda, Arnau, Bogna, Karen, Jan, Christian, Ivan, Michal, Erik, Leonardo, Elsa, Senaida, Jonathan, Alejandro, Ariel, Alexia, Marcus, Flo, Paul, Dani, Tobias, Anthony, Peter, Bobby... I think I'm forgetting some of you I'm sure... to you as well, thanks!

I would like to thank other people in ICFO that made my time extremely enjoyable. Many thanks in particular to Giacomo, Federica, Michela, Mariale, Nadia, Phillip, Mustafa, Boris and many more. Also to the people on admin, HR, KTT and frontdesk for making working here very easy and pleasant.

I am also indebted to my friends out of academia for supporting me during

these years and keeping me sane. Special thanks go to Ignacio, Ricardo from Madrid. From Barcelona I would like to thank Joan, Miry, Abel, Miricornio, Ola, Anna Maria, Rafa, Ester, Jordi, Bea, Pablo, Jesus, Paula, Sheila, Rocio, Ruth, Breo, Txema, Nacho and Monica. I want to greatly thank as well Natalia, Angela, Sara, Sharay and Miriam for their support. Other friends from abroad that supported me during these years include Jo-Anne, Paula, Melissa Sylvia and Katherine. Thank you!

Finalmente, quiero agradecer muy especialmente a mi familia por su amor y apoyo incondicional durante no sólo estos últimos años sino a lo largo de toda mi vida. Muchas gracias a mis padres, que querían ver estas tesis completada incluso más que yo mismo. A mi hermano Salva, autor de la portada, y a mi hermana Fátima. ¡¡Muchas gracias a todos!!

Abstract

Physics has two main ambitions: to predict and to understand. Indeed, physics aims for the prediction of all natural phenomena; for every action, what is the corresponding reaction? Prediction entails modeling the *correlation* between an action, the 'input', and what is subsequently observed, the 'output'.

Understanding, on the other hand, involves developing insightful principles and models that can explain the widest-possible variety of correlations present in nature. Our understanding improves as our principles become more fundamental and typically as the number of them decreases. Complementarily, our understanding also improves as the elegance of our models increases. Remarkably, advances in both prediction and understanding foster our physical intuition and, as a consequence, novel and powerful applications are discovered.

Quantum mechanics is a very successful physical theory both in terms of its predictive power as well as in its wide applicability, from the Higgs boson to all modern electronics. Nonetheless and despite many decades of development, we do not yet have a proper physical intuition of quantum phenomena. I believe that improvements in our understanding of quantum theory will yield better, and more innovative, protocols and vice versa; exploration of new quantum protocols can offer valuable insights on the reality of quantum theory. This dissertation aims at both advancing our understanding and developing novel protocols. This is done through four approaches.

The first one is to study quantum theory within a broad family of theories investigating what principles single out quantum theory out of this family i.e. finding out what is intrinsically quantum. In particular, we study quantum theory within the family of locally quantum theories. We found out that the principle that singles out quantum theory out of this family, thus connecting quantum local and nonlocal structure, is dynamical reversibility. This in turn means that the viability of large scale quantum computing can be based on concrete physical principles that can be experimentally tested at a local level without needing to test millions of qubits simultaneously.

The second approach is to study quantum correlations from a black box perspective thus making as few assumptions as possible. The strategy is to study the completeness of quantum predictions in black box scenarios by benchmarking them against alternative models. Three main results and applications come

out of our study. Firstly, we prove that performing complete amplification of randomness starting from a source of arbitrarily weak randomness - a task that is impossible with classical resources - is indeed possible via nonlocality. This establishes in our opinion the strongest evidence for a truly random event in nature so far. Secondly, we prove that there exist finite events where quantum theory gives predictions as complete as any no-signaling theory (even possibly supra- quantum) can give. This result shows that the completeness of quantum theory is not an asymptotic property. Finally, we prove that maximally nonlocal theories can never be maximally random while quantum theory can, showing a trade-off between the nonlocality of a theory and its randomness capabilities. We also prove that quantum theory is not unique in this respect.

The third approach we follow, closely related to the previous one, is to study quantum correlations in scenarios where some parties have a restriction on the available quantum degrees of freedom while others do not. Interestingly, the future progress of semi-device-independent quantum information, whose promise is to offer robust protocols, depends crucially on our ability to bound the strength of these correlations. Here we provide a full characterization via a complete hierarchy of sets that approximate the target quantum set from the outside. Each set can be in turn characterized using standard numerical techniques termed semidefinite programming relaxations. One of the applications of our work is the ability to certify multidimensional entanglement device-independently.

The fourth approach is to confront quantum theory with computer science principles, exploring the implications of the latter on some of the most fundamental results of quantum theory. In particular, we establish two interesting implications for quantum theory results of raising the Church-Turing thesis to the level of postulate. Firstly, we show how different preparations of the same mixed state, indistinguishable according to the quantum postulates, become distinguishable when prepared computably. Secondly, we identify a new loophole for Bell-like experiments: if some parties in a Bell-like experiment use a computer to decide which measurements to make, the computational resources of an eavesdropper have to be limited to observe a proper violation of non locality.

List of Publications

This thesis is mainly based on the following publications:

- A. Bendersky, G. de la Torre, G. Senno, S. Figueira A. Acín, "Implications of computer science principles for quantum physics". Preprint at arXiv:1312.0265, December 2013.
  Submitted to *Physical Review Letters*

- G. de la Torre, C. Dhara, M. Hoban, G. Prettico, A. Acín, "Maximally nonlocal theories cannot be maximally random".
  Physical Review Letters **114**, 160502 (2015)

- G. de la Torre, C. Dhara, A. Acín, "Certifying the absence of apparent randomness under minimal assumptions".
  Proceedings of 8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)

- C. Dhara*, G. de la Torre*, A. Acín, "Can observed randomness be certified to be fully intrinsic?".
  (*) These authors contributed equally.
  Physical Review Letters **112**, 100402 (2014)

- M. Navascues, G. de la Torre, T. Vertesi, "Characterization of quantum correlations with local dimension constraints and its device-independent applications"
  Physical Review X **4**, 011011 (2014).

- R. Gallego, Ll. Masanes, G. de la Torre, C. Dhara, L. Aolita, T. Acín. "Full randomness from arbitrarily deterministic events"
  Nature Communications **4**, 2654. (2013)

- G. de la Torre, Ll. Masanes, A. J. Short, M. P. Mueller, "Deriving quantum theory from its local structure and reversibility"
  Physical Review Letters **109**, 090403 (2012)

Other works of the author not included in this thesis are:

- L.P. Thinh, G. de la Torre, J-D. Bancal, S. Pironio, V. Scarani, "Randomness in post-selected data"
  In preparation

- A. Bendersky, G. Senno, G. de la Torre, S. Figueira A. Acín, "Determinism and computability imply signaling"
  In preparation

Contents

1. Introduction

1.1. A brief introduction to the area

(Classical) information theory In 1948 Claude Shannon published "A mathematical theory of communication' [Sha48] founding information theory. Few articles can claim to be so influential and to have such big relevance in shaping our modern world. Boosted by the invention of the transistor, information theory brought about the so-called *digital revolution* landing us in a new historical period according to historians: the information age.

A key aspect of Shannon's success was in my view to identify the abstract laws governing information processing on the available physical devices of the time. Its usefulness came from its universality: since the laws he uncovered did not take into account the specifics of those devices, being just in terms of 'information sources', 'channels' or 'information entropy', the results applied equally well to many different practical implementations. The unit of information was defined to be the bit (binary digit) and all the bits, whether represented by a voltage jump or a hole in a paper card, *'were created equal'*.

Physics and information interplay Since the laws were written with such level of abstraction, it was easy to forget one thing: the information laws derived for those bits were assuming implicitly that the underlying physical systems were *classical*, that is, governed by classical physics. However, since the work of Planck, Einstein, Bohr and others, nature had been found to be governed by a more intricate and surprising set of laws termed *quantum mechanics*.

Not just quantitative but also qualitative improvements The application of quantum mechanics to information theory *changed* those laws making them quantitatively better, some examples being super dense coding [BW92] and certain communication complexity tasks [BCMdW10] or algorithms [Sho94]. Even more surprisingly however, quantum mechanics allowed for a plethora of new information processing tasks that previously beyond the realms of possibility. This thesis will present some of these in Chapters 4 and 5. All in all, this revealed a deep lesson to be learned: information is physical [Lan92]. In other

words, the information processing capabilities available to us depend crucially on the laws governing the physical systems we use.

Foundations in physics translate into applications in information tasks From a history of physics perspective, quantum mechanics has become the most successful physical theory in the history of physics with countless experimental corroborations of its validity, ranging from the black body radiation to the recently experimentally discovered Higgs boson. Indeed, on the more practical side, the technological advances that quantum theory precipipated, such as computing devices based on semiconductors and laser technologies, are estimated to be responsible for approximately 30% of the US GDP [TW01].

However, ironically and despite its success, the foundations of quantum mechanics, heavily debated already by some of its founding fathers, constitute still an active area of research. Indeed, the biggest questioning of quantum theory came from Einstein, Podolsky and Rosen who put forward that quantum theory could not be a *complete* theory [EPR35]. The randomness on measurement outcomes that quantum theory described as fundamental would simply come from ignorance on the complete set of so far hidden variables describing the state and measurement process. John Bell advanced greatly the issue in his 1964 seminal paper [Bel64] by showing that, under two very fundamental assumptions; that no signal can be instantaneously transmitted and that those performing an experiment have freedom of choice, such hidden variables could not exist. As counter-intuitive as it may sound and modulo the said assumptions, the information of which measurement outcome will appear *did not 'exist' before the measurement*. Correlations for which no hidden-variable explanation exists are termed nonlocal. As we shall see, quantum theory is said to be nonlocal as it permits nonlocal correlations to be established.

Far from closing the debate, this incredible result opened the 'pandora box' of both fundamental questions and information tasks, solutions to the first being readily translated into proposals for the latter. More concretely, questions related to the completeness of quantum theory holds strong connections with randomness generation tasks and cryptography as we shall see during the thesis. All in all, the main conclusion can be that not only information is physical, but *physics is information too*.

1.2. Motivation and main contributions

This thesis lies at the intersection between physics and information, between foundational questions and information protocols. The thesis contributes to

different questions in the areas of axiomatic reconstructions of quantum theory, semi-device independent quantum information, device-independent randomness certification tasks and quantum randomness and computability.

1.2.1. Axiomatic reconstructions of quantum theory

Quantum theory emerges from its local structure plus reversible dynamics

Most physical theories can be derived from a few axioms that accept a non-technical explanation in physical terms. For example, special relativity can be derived from two axioms namely the 'principle of relativity' that states that the laws of physics are the same in all inertial reference frames, and the 'invariance of light-speed', that says that the speed of light in free space has the same value in all inertial reference frames. One may think that the existence of a formulation of the theory in terms of few, simple and physically motivated axioms points to a good understanding of the theory. Arguably, such axioms can foster physical intuition.

Quantum theory, however, is commonly presented via a set of heavily mathematical axioms. To present quantum theory one usually needs to consider positive operators acting on a Hilbert space, unitary matrices, the Born rule etc. The quest to find physical axioms from which to derive quantum theory is as old as the theory but in the past few years a wave of works have appeared. Most of them are based on the seminal paper by Lucien Hardy [Har01], from which Jonathan Barrett derived a probabilistic and operational framework where different theories could be defined based on different axioms [Bar07]. This framework is called Generalized Probabilistic Theories (GPTs). From then on, a number of works have used this framework to either reconstruct quantum theory from a set of proposed axioms (more physically or informationally oriented) or to explore alternative theories. Assuming only a part of quantum theory and still be able to derive certain results, previously thought to be purely quantum, helps us to understand better what is genuinely quantum.

Recently the class of physical theories with the same local structure as quantum theory, but a potentially different global structure has been introduced. It has previously been shown [AAC$^+$10, BBB$^+$10] that any bipartite correlations generated by such a theory must be simulatable in quantum theory making those theories hence undistinguishable in a device-independent manner. However, it was also shown in [AAC$^+$10] that this does not hold for tripartite correlations. What is then the missing ingredient that singles out quantum theory from this family of theories?

Contributions

Our contribution is to identify an additional constraint on the space of locally quantum theories that allows us to recover the global quantum structure. That condition is that the dynamics should be reversible. Specifically, in the particular case in which the local systems are identical qubits, we show that any theory admitting at least one continuous reversible interaction must be identical to quantum theory.

Another implication of our work is on the viability of large scale quantum computing. Some authors have argued that quantum theory could experience a fundamental breakdown once a particular threshold in the number of qubits has been achieved. Our results make this possibility more unlikely since, if so, one of our assumptions must be violated. That is, either local quantumness, or continuous reversibility, or local tomography, must fail. Local tomography establishes that states of a composite system are completely characterized within the theory by the statistics on the subsystems. These violations could then be addressed experimentally, without necessarily having to test this on a macroscopic number of qubits.

1.2.2. Device-independent randomness certification tasks

Full randomness amplification is possible

Do completely unpredictable events exist? Of course this fundamental question is of big interest for mutiple thinkers from diverse areas ranging from neuroscientists to philosophers, but what is the biggest evidence we have from the physics perspective? To begin with, we know that classical physics excludes fundamental randomness, leaving room only for randomness resulting from ignorance or lack of control. While quantum theory makes probabilistic predictions, this does not imply that Nature is random, as randomness should be certified without relying on the complete structure of the physical theory being used.

Bell tests approach the question from this perspective, making very few assumptions that do not involve the mathematical structure of the theory. However, they require prior perfect randomness to choose the settings, falling into a circular reasoning. Remarkably, a Bell test that generates perfect random bits from bits possessing high -but less than perfect- randomness was recently obtained in [CR12b]. Yet, the main question remained open: do sources of arbitrarily weak initial randomness suffice to certify perfect randomness?

Contributions

We answer the previous question by providing the first protocol attaining *full randomness amplification*, that is, a protocol that using an arbitrarily weak source of random bits produces arbitrarily perfect ones. It works by feeding the initial bits into a multipartite Bell test and only assumes the impossibility of instantaneous communication. Given that the result is derived under minimal assumptions, we argue that it constitutes the biggest evidence so far for the existence of a purely random event. Interestingly, the level of security of our result when considered from a cryptographic perspective is the highest possible (universal composability) which makes it ideal as an initial module to any other device-independent protocol.

Completeness of finite-size quantum events

In general, any observed random process includes two qualitatively different forms of randomness: apparent randomness, which results both from ignorance or lack of control of degrees of freedom in the system, and intrinsic randomness, which is not ascribable to any such cause. While classical systems only possess the first kind of randomness, quantum systems may exhibit some intrinsic randomness as they have been shown to be capable of producing non-local correlations. However, quantum correlations are not the most nonlocal correlations respecting the no-signalling principle as famously showed by Popescu and Rohrlich [PR94]. This implies that quantum events have typically both forms of randomness intertwined. Even for our previous result showing full randomness amplification, for any finite-size scenarios there is a gap between the observed randomness and the intrinsic randomness. Could it be the case that for all finite scenarios there always exists a gap between the randomness that we observe and that which we can certify? What if completeness of quantum theory is only an asymptotic property and, therefore, quantum predictions can be completed in any finite setup?

Contributions

We answer the previous question by identifying the first family of quantum processes whose observed randomness can be proven to be fully intrinsic.

These results are derived under minimal assumptions: the validity of the no-signalling principle and an arbitrary (but not absolute) lack of freedom of choice. This implies that these events cannot be further completed, in the sense that a theory giving better predictions for these events should be either signalling or have no freedom of choice.

Our results prove that quantum predictions cannot be completed already in simple finite scenarios, for instance of three parties performing two dichotomic measurements. Moreover, the observed randomness tends to a perfect random bit when increasing the number of parties, thus defining an explicit process attaining full randomness amplification.

Maximally nonlocal theories cannot be maximally random

Correlations that violate a Bell Inequality are said to be nonlocal, i.e. they do not admit a local and deterministic explanation. Great effort has been devoted to study how the amount of nonlocality (as measured by a Bell inequality violation) serves to quantify the amount of randomness present in observed correlations. In this work we reverse this research program and ask what the randomness certification capabilities of a theory tells us about the nonlocality of that theory.

In particular, we are interested in the following questions. Why is nature governed by a nonlocality-restricted theory such as quantum theory? What is the relationship between nonlocality and randomness within quantum theory? Does more nonlocality in a theory imply more randomness? And more specifically, are maximally nonlocal theories also maximally random? Is the bounded nonlocality of quantum theory enough for maximal randomness certification? Are there other theories with similar properties?

Contributions

We contribute to the given research program by proving that, contrary to initial intuition, maximal randomness certification cannot occur in maximally nonlocal theories. Indeed, there is always some non-trivial predictability in an experiment with no nonlocality restrictions.

We go on and show that quantum theory, in contrast, permits certification of maximal randomness in all dichotomic scenarios. We hence address the question of whether quantum theory is optimal for randomness, i.e. is it the most nonlocal theory that allows maximal randomness certification?

We answer this question in the negative by identifying a larger-than-quantum set of correlations capable of this feat. We think these results will be relevant not only to understand quantum mechanics' fundamental features, but also put fundamental restrictions on device-independent protocols based on the no-signaling principle.

1.2.3. Semi-device independent quantum information

Characterizing quantum correlations with bounded dimensions and its applications

The nascent field of device-independent quantum information has grown rapidly during the last decade, establishing many important results such as quantum cryptography, full randomness amplification or self-testing. However, the requirements that it is based on turn out to be too demanding for many applications. A possibility is to relax them by identifying situations where one can assume a bound on the effective Hilbert space dimension where the state of some physical systems live on. This approach is called semi-device independent quantum information. Apart from the fundamental interest of characterizing quantum correlations with dimension constraints, the future progress of semi-device independent quantum information depends crucially on our ability to bound the strength of the nonlocal correlations achievable with finite dimensional quantum resources.

Contributions

Our contribution is to characterize quantum nonlocality under local dimension constraints via a complete hierarchy of sets that approximate our target set from the outside. These sets can be characterized using standard numerical techniques termed semidefinite programming relaxations.

When applied to bipartite cases, the hierarchy allowed us to give non-trivial bounds to scenarios up to four measurement settings on one side and twelve in the other in a normal desktop.

In the tripartite case, we applied it to certify three-dimensional entanglement in a device-independent way. We did this deriving a Bell-type inequality that can only be violated when each of the three parties has local dimension greater than two. Finally, we show how the new method can be trivially modified to detect non-separable measurements in two-qubit scenarios.

1.2.4. Quantum randomness and computability

Implications of Church-Turing thesis for quantum physics

The Church-Turing thesis is one of the pillars of computer science; it postulates that every classical system has equivalent computability power to the so-called Turing machine. While this thesis is crucial for our understanding of computing devices, its implications in other scientific fields have hardly been explored. In

particular, what are the implications of raising the Church-Turing thesis to the level of postulate?

Contributions

We address that question by starting this research programme in the context of quantum physics and show that computer science laws have profound implications for some of the most fundamental results of the theory. We first show how they question our knowledge on what a mixed quantum state is, as we identify situations in which ensembles of quantum states defining the same mixed state, indistinguishable according to the quantum postulates, do become distinguishable when prepared by a computer.

We also show a new loophole for Bell-like experiments: if some of the parties in a Bell-like experiment use a computer to decide which measurements to make, then the computational resources of an eavesdropper have to be limited in order to have a proper observation of nonlocality. We believe that our work opens a new direction in the search for a framework unifying computer science and quantum physics.

2. Preliminaries

In this Chapter we introduce all the different concepts and tools that will be needed during the thesis. Common to all parts is the view of physics from a *black box* perspective. This perspective is particularly amenable to an information-theoretic assessment of the power of the theory to perform any information task such as computation or cryptography, without taking into account any details of all possible implementations.

2.1. Generalized probabilistic theories (GPTs)

The framework of generalized probabilistic theories, also known as convex operational theories, constitutes a simple mathematical framework that allows one to write down different probabilistic theories from a black box perspective. Interestingly it subsumes both classical and quantum theory as particular instances of GPTs while allowing for a rich universe of different theories to be defined. We will use this framework to study the family of locally quantum theories in Chapter 3.

The said framework assumes that, whatever the physical theory might be describing a given physical systems, there exists a *classical* macroscopic level in which it makes sense to talk about experimentalists that build experimental setups. Those experimentalists must be able to perform preparations, throw biased coins to prepare mixtures, observe macroscopic measurement outcomes (such as detector clicks) and write down data tables so as to compute relative frequencies.

To introduce the framework, we will make use of Figure 2.1 which describes in an abstract way what a general physical experiment is.

Imagine a laboratory that has three type of devices each with different knobs defining different configurations: a preparation device, a transformation device and a measurement device. The preparation device will prepare the *system* for each configuration of the knob in a particular *state* ψ. The transformation device, when placed in a particular configuration, will change the state of the system which will posibly translate into a change in the probability of some measurement outcome x to occur. Which measurement to perform can again be chosen by changing a knob in the measurement device.

The GPT framework was introduced in modern form by Barrett in [Bar07] by formalizing the background assumptions and implicit framework used in the important article by Hardy [Har01] that shows how to reconstruct quantum theory from five simple axioms. However, this framework points back originally to Mackey [Mac63] and has been rediscovered many times. For an historical account see [Har13].

More recently, after the work of Barrett, a great number of works have appeared using the GPT framework to study how the pieces of quantum theory fit together. The fruits of these studies can be divided in three types. First, articles were results commonly thought to be purely quantum are proved to hold for a broad family of theories, based in much fewer assumptions. Examples of these are the no-cloning theorem [Bar07], no-broadcasting theorem [BBLW07] or the existence of nonlocality [Bar07] or the monogamy of correlations [MAG06]. Second, reconstructions of quantum theory from a few physical or information assumptions ie. [DB11, MM11, CDP11] and also our contribution [dlTMSM12]. Third and last, works that study alternatives to quantum theory, usually sharing some structure with quantum theory but dropping some assumptions ie. [HW10, GMCD10, JGBB11, MOD12].

In what follows we will make a brief introduction to the main ingredients of the framework needed for Chapter 3, stating some of the basic results without a proof. We will follow closely the introduction to GPTs by [MM11] as this best suits the purpose of this thesis. For a longer and more detailed explanation the reader is refered to the articles [Bar07, MM11] or [CDP11] for a slightly different presentation style of the same framework.

2.1.1. Systems and states

We will say two physical systems are in the same state ψ if all outcome probabilities of all possible measurements performed under the same transformations are the same. We will imagine that systems come in different *types*. For example, in quantum theory each type can be thought of being parametrized by the Hilbert space dimension. The probability of a measurement outcome x will be denoted with $p(x)$ and will be associated with one possibility of a binary measurement (something happening or not). The complementary event $\overline{x}$ (not happening) can be derived from the previous one by $p(\overline{x}) = 1 - p(x)$. Notice that a measurement with more than two outcomes can always be rewritten as several measurements of two outcomes and hence there is not lack of generality.

Within a particular type of system, we will assume that the outcome probabilities of a *finite* set of measurements is sufficient to derive outcome probabilities for all the other possible measurements. This is a crucial assumption that

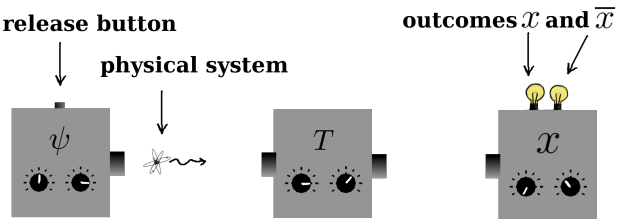

Figure 2.1.: General experimental set up. Every experiment is performed in three steps through three type of devices. First the preparation device equipped with a knob to choose among several states. After the release botton is pressed and the system prepared, it enters into the transformation device where again, a transformation parametrized by another knob takes place. Finally the systems enters the measurement device whose measurement is parametrized through a third knob and gives a dichotomic outcome (x or $\overline{x}$), indicated by the corresponding light.

excludes the treatment of systems of 'infinite capacity' or continuous-varible systems. Although those could be relevant for some physical situations, we will not treat them here. It is arguable, however, whether a system of *infinite* capacity is physical or just an approximation of a very big but finite capacity system. We will however leave this interesting discussion aside and concentrate on finite systems. For example, in quantum theory, for a system of 'the second type' (a system acting on a Hilbert space dimension two, that is a $\frac{1}{2}$-spin particle) it is enough to know the outcome probabilities of the three Pauli observables $\{\langle\sigma_x\rangle, \langle\sigma_y\rangle, \langle\sigma_z\rangle\}$ to reconstruct the state of the system ρ, making tomography of the state. One can later derive the probability of any other measurement outcome. Indeed, once the state is known, we can derive the probability of outcome $\pm$ for a spin measurement along any other direction given by a unit vector $\hat{n}$: $p(\pm) = \mathrm{tr}(\rho \cdot \frac{\mathbb{I}\pm\sigma_{\hat{n}}}{2})$ with $\sigma_{\hat{n}} = \vec{\sigma}\cdot\hat{n}$ and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. We will call one of the smallest of such sets of measurements *fiducial* and will use them to describe the state of the system. Let us denote by $x_1, x_2, \ldots, x_d$ these measurements, the state will then be described by

$$\psi = \begin{bmatrix} 1 \\ p(x_1) \\ \vdots \\ p(x_d) \end{bmatrix} = \begin{bmatrix} \psi^0 \\ \psi^1 \\ \vdots \\ \psi^d \end{bmatrix} \in \mathcal{S} \subset \mathbb{R}^{d+1}. \tag{2.0}$$

The set of fiducial measurements need not be unique (for example in QT the state can be defined by giving the probability of outcomes in three linearly independent directions). The first component of the state $\psi_0 = 1$ is redundant but will make the notation simpler when dealing with composite systems, leading naturally to the use of tensor products.

Now we will introduce a couple of properties of the state space $\mathcal{S}$. First, note that it is convex [Roc70] since, if ψ_1, ψ_2 belong to $\mathcal{S}$, then one can prepare the convex mixture $p\psi_1 + (1-p)\psi_2$ by preparing ψ_1 with probability p and ψ_2 with probability $1-p$. Second, it is bounded, because each entry is a probability in the interval $[0, 1]$. The affine dimension of the space must be d since otherwise at least one of the components would be functionally related to the others and hence redundant, contradicting what we defined to be a *fiducial set of measurements*.

Most importantly, we will assume the set $\mathcal{S}$ to be topologically closed. The reason being that, although this assumption has no physical relevance as we shall see, it will make the mathematical treatment simpler and more elegant allowing us to use the machinery of compact convex sets. Physically it does not make any difference since, given that the states are defined by probabilities for certain outcomes, there is no physical difference between a state ψ and an arbitrarily good approximation to it. That is, the states in the topological closure are indistinguishable from some states in the interior. In that way we can include the closure of the states we can prepare to be part of the state space $\mathcal{S}$ making it closed. Now, since we are assuming the state vectors to be finite-dimensional, bounded, and the set spaces $\mathcal{S}$ topologically closed, we can think of the state spaces $\mathcal{S}$ as *compact convex sets* [Roc70].

Pure states of the state space $\mathcal{S}$ are hence *extremal points* of the convex set, ie. those that cannot be written as a convex mixture $\psi \neq p\psi_1 + (1-p)\psi_2$ for any two different states ψ_1,ψ_2 and $0 < p < 1$. Since $\mathcal{S}$ is compact and convex, all states are mixtures of pure states [Roc70].

2.1.2. Measurements

The probability of obtaining outcome x when a particular measurement has been performed on a state $\psi \in \mathcal{S}$ is given the function $p(x) = \mu_x(\psi)$.

Importantly, measurements must act on the states in a linear fashion. We will later see that transformations also act linearly. One could think that we are making an assumption in order to restrict ourselves to linear theories and hence excluding non-linear evolution or measurements. However, this is not the case. The reason for the linearity is derived very naturally from the convexity of the state space. To further explain why, imagine the following situation:

experimentalist A prepares the mixture state $\psi = p\psi_1 + (1-p)\psi_2$ by flipping a biased coin which gives heads with probability p, after which he prepares ψ_1 and tails with probability $1-p$ followed by a preparation of ψ_2. The state prepared in this way is passed on to a experimentalist B which performs a measurement $\mu_x(\psi)$. If experimentalist A does not inform B about what the outcome of his coin flip, the probability of obtaining an outcome x on the state must be $p(x) = \mu_x(p\psi_1 + (1-p)\psi_2)$. If on the other hand he tells experimentalist B the outcome of the coin flip before he subjects the system to the same measurement the outcome probability must be $p\mu_x(\psi_1) + (1-p)\mu_x(\psi_2)$. It is clear that the measurement cannot be affected by this information, that is, the probability $p(x)$ cannot change and both quantities must be equal.

Mathematically, this means that μ_x is affine on $\mathcal{S}$, as it preserves barycenters. Moreover, the redundant component $\psi_0 = 1$ of our 'augmented' vector state allow us to treat this function as a linear map $\mu_x : \mathbb{R}^{d+1} \to \mathbb{R}$ [Bar07].

Associated with the probability of any particular outcome x there will be an *effect* μ_x with two properties. First, being linear an second, giving rise to a valid probability, that is, $\mu_x : \mathbb{R}^{d+1} \to \mathbb{R}$ and $\mu_x(\psi) \in [0,1]$ for all states $\psi \in \mathcal{S}$. The converse however is not true as there might be mathematically well defined linear maps which are however not valid measurements in the theory as those are banned for some reason. This is analogous to the case of superselection rules that impose axiomatic restrictions that forbid certain states or measurements for some physical symmetry argument or conservation law. For example, a superselection rule for electric charge forbids the preparation of a coherent superposition of different charge eigenstates, see for instance [Haa92].

2.1.3. Transformations

We have seen that systems can be in particular states from the state space and how to measure them. A GPT theory however is specified by giving, for every type of system, the state space, the set of measurements and the set of transformations. A transformation brings the system from one state to another, that is T is a map $T : \mathcal{S} \to \mathcal{S}$. As promised an analogously to the measurements, it must act affinely on the state space

$$T(p\psi_1 + (1-p)\psi_2) = pT(\psi_1) + (1-p)T(\psi_2)$$

as the transformation device cannot act differently depending on whether the outcome of coin flip during the preparation is passed on to the experimentalist operating the device or not. Again the redundant component ψ_0 allow us to make T a linear map $T : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ [Bar07]

For the purpose of this thesis there will be a type of transformations of the outmost importance: *reversible transformations*. A transformation T is reversible if the inverse transformation T^{-1} both exists and belongs to the set of allowed transformations by the theory, ie. is not banned from the set through a super-selection-like rule. The allowed set of reversible transformations of a theory forms a group $\mathcal{G}$ since clearly they comply with the group axioms (associativity and the existence of an identity and inverse group element). Analogously to what we said for the state space, we can assume that the group $\mathcal{G}$ is topologically closed without changing the predictions of the theory. Since the group $\mathcal{G}$ acting on $\mathcal{S}$ maps the bounded state space into itself, it must be a bounded set of matrices. Moreover, it is shown in [Bak02] that all compact matrix groups are Lie groups, in the general sense, where a finite group is also a Lie group with trivial Lie algebra (see [Bak02]). Reference [MM11] elaborates on this.

In a nutshell, Lie groups are groups which are also smooth and real manifolds, that is, smooth manifolds obeying the group properties and the additional requirement that the group operations are differentiable. Moreover to every Lie group $\mathcal{G}$ we can associate a vector space tangent of that Lie group at the identity called Lie algebra $\mathfrak{g}$. Informally speaking the Lie algebra are elements of the group that are *infinitesimally close* to the identity and hence describe how the group looks *locally*. For connected groups ie. where any two elements can be connected by a smooth curve lying entirely within the group, *each* element of the group is related to an element of the algebra by exponentiation: for each $H \in \mathcal{G}$ there is an element $X \in \mathfrak{g}$ such that $H = \mathrm{e}^X$.

Lie algebras also come with a non-associate multiplication $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called Lie bracket that obeys the abstract laws of the commutators, such as bilinearity or the Jacobi identity (see [Bak02] for a formal definition).

The key insight is that, for connected Lie groups one can completely characterize the group, which is a geometrical object, through its corresponding algebra which is a linear vector space that can be easier to work with.

Examples of connected Lie groups are the three-dimensional rotation group SO(3) and the unitary group U(n). Their corresponding Lie algebras are denoted by $\mathfrak{so}(3)$ and $\mathfrak{u}(n)$. The Lie algebra of the rotation (unitary) group are skew-symmetric (hermitian) matrices.

2.1.4. Composite systems

For most of the interesting information-theoretic questions, we need to have a clear notion of how several systems combine into a global system. Importantly, we should have a notion of *local* operations such as fiducial measurements which

apply to subsystems alone without modifying the state of the other subsystems. We will say hence that two systems A and B constitute a composite system AB (and not just one system) if a measurement for A together with a measurement for B constitute a measurement for AB so that the temporal order in which we perform the measurements does not make a difference (local operations commute). In other words, if x and y are the measurement outcomes on systems A and B, the probability of those outcomes ocurring $p(x, y)$ should not depend on the temporal order in which the systems were measured. This implies we can think of the systems as being measured simultaneously without any ambiguity.

This notion of composite systems with subsystems in well-defined marginal states can be imposed through the so-called *no-signaling principle*:

Definition 2.1 (No-signaling principle)**.** *Consider two physical systems A and B. A theory describing those systems obeys the no-signaling principle if from operations on system A is not possible to extract information about which operation was performed on system B.*

For example, if we name by $x_1, \ldots, x_{d_A}$ the fiducial measurements of system A and by $y_1, \ldots, y_{d_B}$ the fiducial measurements of system B, the no-signaling conditions read

$$\begin{array}{rcl} p(x_i) & = & p(x_i, y_j) + p(x_i, \overline{y}_j), \\ p(y_i) & = & p(x_j, y_i) + p(\overline{x}_j, y_i), \end{array} \tag{2.1}$$

for all i, j.

Notice that stating the conditions for the fiducial measurements is enough since the probabilities for other measurements are linear combinations of the former. When we introduce the device-independent scenario we will see the same principle at work in a slightly different context.

An additional assumption that is usually made at this level is called local tomography [MM11], also known as global state assumption [Bar07] or local discriminability [CDP11]. It states that

Definition 2.2 (Local tomography)**.** *A theory respects the local tomography principle if the state of a composite system is completely characterized within the theory by the statistics of measurements on the subsystems. That is, $p(x, y)$.*

Although some work has been done studying GPTs were this assumption does not hold (see for instance [HW10]. There local tomography is shown not to hold on real-vector-space quantum theory) most works assume it as a natural requirement. We will make this assumption in Chapter 3 as well. Importantly,

local tomography (2.2) together with the no-signaling principle (2.1) imply that states in the composite system AB can be represented on the tensor-product space [Bar07] as

$$\psi_{AB} = \begin{pmatrix} 1 \\ \vdots \\ p(x_i) \\ \vdots \\ p(y_j) \\ \vdots \\ p(x_i, y_j) \\ \vdots \end{pmatrix} \in \mathcal{S}_{AB} \subset \mathbb{R}^{d_A+1} \otimes \mathbb{R}^{d_B+1}. \tag{2.1}$$

Likewise, the probability for the measurement outcomes x, y in subsystems A and B will be given by

$$p(x, y) = (\mu_x \otimes \mu_y)(\psi_{AB}) \tag{2.1}$$

where μ_x and μ_y are the measurement effects representing outcomes x, y, ie. $p(x) = \mu_x(\psi_A)$ and $p(y) = \mu_y(\psi_B)$. Trasformations on each subsystem act in the global system as $\psi_{AB} \to (T_A \otimes T_B)(\psi_{AB})$, where T_A and T_B represent transformations on the reduced states ψ_A and ψ_B. Indeed, these reduced states are obtained from the global state ψ_{AB} simply by choosing the right components. Reduced states can also be defined by saying ψ_A is the reduced state of subsystem A in the global system AB iff $\mu_A(\psi_A) = (\mu_A \otimes \mathbb{I})(\psi_{AB})$ for all effects μ_A and where $\mathbb{I}(\psi_B) = \psi_B^0 = 1$ is the unit effect. The reduced states ψ_A, ψ_B must belong to the subsystem's state space $\mathcal{S}_A, \mathcal{S}_B$ and any state in $\mathcal{S}_A, \mathcal{S}_B$ must be the reduction from a state in the global state space $\mathcal{S}_{AB}$. This implies that all product states $\psi_A \otimes \psi_B$ belong to $\mathcal{S}_{AB}$ [Har09a]. And likewise for the measurements and transformations: all local tensor products of measurements and transformations are allowed on system AB. We can use this to define what is a entangled state (or entangling measurement/transformation) in the general context of GPTs.

Definition 2.3 (Entanglement in GPTs)**.** *We say a composite system AB in a state ψ_{AB} is entangled iff it cannot be written as a convex (probabilistic) combination of product states*

$$\psi_{AB} \neq \sum_i p_i \psi_A^i \otimes \psi_B^i \tag{2.1}$$

with $p_i \geq 0$ and $\sum_i p_i = 1$.

Similar definitions apply to measurements or transformations that cannot be written as probabilistic mixtures of tensor-products of local measurements or transformations. When the GPT is quantum theory this definition recovers the usual definition of quantum entanglement, entangling dynamics and measurements. It is hence interesting to see that these properties are not unique to quantum theory but are quite natural in this broader and more general framework.

Notice that, according to the previous considerations, given two fixed state spaces $\mathcal{S}_A$ and $\mathcal{S}_B$ there are some constraints that the composite system $\mathcal{S}_{AB}$ must fulfil. However, there is still a lot of freedom that allows one to define many different theories, some of which may have different information-processing capabilities (such as different Bell inequality violations) as we shall see. Of course the definition of composite system was made for two-systems but it can be extended recursively to an arbitrary number of systems.

2.1.5. Equivalent theories

Imagine two theories described by their state spaces, measurements and transformations labelled by $\mathcal{S}, \mu, T$ and $\mathcal{S}', \mu', T'$ respectively. Moreover, imagine that there exists an invertible affine map $\mathcal{L}$ that connects the first theory with the second acting in the following way: states are transformed as $\psi \to \mathcal{L}(\psi)$. All effects in $\mathcal{S}$ are transformed like $\mu \to \mu \circ \mathcal{L}^{-1}$ and the transformations like $T \to \mathcal{L} \circ T \circ \mathcal{L}^{-1}$. Notice that both theories give the same predictions, that is, the same measurement outcome probabilities:

$$\begin{array}{lcl} p'(x) & = & \mu'(T'(\psi')) = (\mu \circ \mathcal{L}^{-1})(\mathcal{L} \circ T \circ \mathcal{L}^{-1}(\mathcal{L}(\psi)) = \\ \mu(T(\psi)) & = & p(x) \end{array} \tag{2.1}$$

That is, $p'(x) = p(x)$. We will hence say that these two theories are indeed *equivalent* in the sense that they are just different representations of the same theory. The representation carries no physical significance. For example, in the standard formalism of quantum theory, states are represented as complex-valued density matrices, but that is an equivalent representation to the one in terms of probabilities described in Section 2.1.1 and there exists a map $\mathcal{L}$ connecting both representations. In the same way, one can equivalently describe the state of a qubit with different representations such as the density matrix or the Bloch vector.

2.1.6. Instances of GPTs

In this subsection we will describe several GPTs or families of GPTs that are of special relevance for this thesis. Recall that a GPT is uniquely determined by specifying a representation of the state space $\mathcal{S}$, the set of allowed measurements on $\mathcal{S}$ denoted by μ and set of allowed transformations $\mathcal{T}$.

Quantum theory (QT)

It is a particular GPT which can be defined by specifying that the state space $\mathcal{S}_c$ of $c-$level quantum systems is equivalent to the set of complex $c \times c$-matrices ρ such that $\rho \geq 0$ and $\mathrm{tr}(\rho) = 1$. The dimension of this set is $d_c = c^2 - 1$ (one less because of the $\mathrm{tr}(\rho) = 1$ condition) and pure states are matrices of rank one. The effects on $\mathcal{S}_c$ act like $\mu(\rho) = \mathrm{tr}(M\rho)$ where M is a $c \times c$ complex and positive semidefinite matrix such that $0 \leq M \leq \mathbb{I}$. Transformations from the allowed set $\mathcal{T}$ are all reversible and act like $\rho \to U\rho U^\dagger$ with $U \in SU(c)$, that is, $c \times c$ unitary matrices with unit determinant to preserve normalization.

Importantly, when two systems A and B living in $\mathcal{S}_{c_A}$ and $\mathcal{S}_{c_B}$ respectively combine, the global state space $\mathcal{S}_{AB}$ is indeed equivalent to $\mathcal{S}_{c_{AB}}$ where $c_{AB} = c_A \cdot c_B$. The allowed set of transformations and measurements are those acting on $\mathcal{S}_{c_{AB}}$.

Locally quantum theories (LQT)

The following family of GPTs will be of special relevance for this thesis, specifically for Chapter 3. It can be defined by the following: The local state spaces, measurements and transformations are equivalent to those of QT, given in the previous description by $\mathcal{S}_c$, $\mathcal{T}$ and μ. However, the last condition of the QT description about how to combine several systems does not neccesarily hold, ie. $\mathcal{S}_{AB} \neq \mathcal{S}_{c_A \times c_B}$. This family of theories will allow us to separate the *local* and *nonlocal* aspects of quantum theory, where the first holds but not the latter. The global state spaces in this family are still constrained by the no-signaling principle (2.1) and local tomography (2.2) and hence share many properties with standard quantum theory. This further freedom allows nevertheless for a family of GPTs some of which permit for, for instance, supra-quantum Bell inequality violations [AAC$^+$10].

QT can be understood as a particular instance within this family of theories. In Chapter 3 we will study what makes QT unique in this family of theories, shedding light on why would nature be quantum mechanical.

Classical probability theory (CPT)

It is another relevant instance of a GPT which can be analogously defined. The state space $\mathcal{S}_c$ is that of $c-$outcome probability distributions $[p(1), \ldots, p(c)]$ with dimension $d_c = c-1$ (again, one less because of normalization). This means in geometrical terms that the state space $\mathcal{S}_c$ is a simplex . Pure states are given by deterministic distributions $p(x) = \delta_{x,y}$ with $y \in [1, \ldots c]$. Measurements are functions of the $c-$outcome complete measurement that distinguishes c states given by $\mu_a(\psi) = p(a)$ for $a = 1, \ldots, c$. Interestingly, CPT can be thought of a restriction of QT where the allowed states must be diagonal (no coherences). We will hence say then that CPT is embedded in QT.

Remarkably, all physical systems which can be described purely through classical physical laws, such as classical mechanics, electromagnetism or general relativity, are fully described by the CPT when treated in the framework of generalized probabilistic theories. This explains, loosely speaking, why all classical physics have the same information processing power (although the information processing power of classical continuous systems would deserve a more subtle discussion). Indeed, that is why classical information theory, started with Shannon's seminal work [Sha48], was thought to be *physics-independent*, as it applied equally well to all physical systems of that time.

Generalized no-signaling theories (GNS)

This family of GPTs introduced in [Bar07] will be of special relevance for Chapter 6. The state space for these theories contains, by definition, all states producing correlations that satisfy the no-signaling constraints (recall 2.1). These state spaces contain finitely many extreme points and some of them are proved to be significantly more non-local than what quantum theory allows [PR94]. Effects in these theories are products of local effects. With respect to the transformations, different theories can be constructed depending on what type of transformations are considered. When restricted solely to reversible dynamics, the transformations consist only of relabelings of local measurements, outcomes and permutations of subsystems (or combinations of the latter) as proved in [GMCD10]. In Chapter 6 we will study the question of what distinguishes these GNS theories from quantum theory and provide some answer in term of randomness certification capabilities. We will shed some light on why nature is not as non-local as it could be while still respecting the no-signaling principle.

2.2. Device-independent framework

The device-independent framework shares with the previous Section the approach to physics from a black box perspective. However, it differs in the assumptions it makes which defines what these two frameworks are useful for.

The interest of the GPT framework is to study the capabilities and structure of the effective (probabilistic) theory describing some given physical systems and their interaction. As such, the experiments described in Section 2.1 through Figure 2.1 are idealized experiments and, very importantly, *non adversarial*. This implies importantly that within this framework one can perform tomography of a particular state ψ by repeating the process of prepare and measure and gathering statistics. Implicitly there is the assumption that every time we press the release button the same state is prepared. Moreover, we assume that every time one presses a measurement and transformation button on a certain configuration, the same measurement and transformation is performed irrespective of the state we are making tomography of. These may not be true in a typical adversial scenario.

The object of interest of the device-independent framework is, on the other hand, the correlations that can be established between parties irrespective of the exact state, measurement and transformation that gave rise to those correlations. In the device-independent framework weaker assumptions are made, hence making it ideal for designing robust information protocols that do not depend on an accurate description of the physical model and parameters describing the inner working of the black boxes. This makes it very convenient to assess the security of those information protocols. Examples of those protocols are those solving tasks such as key distribution, randomness expansion, randomness amplification or self-testing (see [BCP$^+$14] for references on these and many more protocols).

The device-independent framework considers a number of parties, say N, situated in different locations and that possess black box measurement devices. These devices will admit a number of configurations, say M, that can be chosen by changing a knob, pressing a button or by some other macroscopic means. The possible number of macroscopic outcomes for each measurement will be d. These outcomes can be encoded in a light switch, a pointer, etc.

A typical device-independent protocol has many rounds that are repeated. Let us describe the first one of them: a source situated at a distant location will prepare a composite physical system composed of N different subsystems in an unknown state. Each of this subsystems will be sent to one of the N parties denoted by $A_1, A_2, \ldots, A_N$. Upon reception in each party's laboratories, the subsystem will enter the black box measurement device. Party i will then select

at random what measurement to make by changing the corresponding knob. This choice will be labelled by the variable $x_i \in \{0, \ldots, M-1\}$. The outcome of the measurement will be labelled by $a_i \in \{0, \ldots, d-1\}$. Both choices of input and output will be written down in each party's data table. A device-independent scenario is hence defined by the triple of number of parties, inputs and outputs (N, M, d).

The pair of vectors of classical variables $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{a} = (a_1, \ldots, a_N)$ collect the inputs and outputs of that round of the experiment. The probability with which these vectors of inputs and outputs will appear are governed by a so far unknown probability distribution $P(\mathbf{a}|\mathbf{x}) = P(a_1, \ldots a_N|x_1, \ldots, x_N)$ which however, is assumed to exist.

For simplicty and pedagogical reasons, let us make two preliminary assumptions: First, let us assume the so called *i.i.d assumption*, where i.i.d stands for independent and identically distributed systems. This assumption implies that the source is making the same preparation in every round of the experiment. Another assumption that is typically implied in the i.i.d scenario is that the devices behave as if they had no memory of previous rounds. Let us denote by N_{tot} the total number of rounds in the experiment. It is clear then that the observed relative frequencies $n(\mathbf{a}|\mathbf{x})/\sum_{\mathbf{a}} n(\mathbf{a}|\mathbf{x})$ where $n(\mathbf{a}|\mathbf{x})$ enumerates the number of rounds were outcomes $\mathbf{a}$ happened given that the measurement inputs $\mathbf{x}$ were chosen will tend to the previously mentioned probabililty distribution $P(\mathbf{a}|\mathbf{x})$ as N_{tot} grows. More specifically:

$$P(\mathbf{a}|\mathbf{x}) = \lim_{N_{\text{tot}} \to \infty} \frac{n(\mathbf{a}|\mathbf{x})}{\sum_{\mathbf{a}} n(\mathbf{a}|\mathbf{x})} \tag{2.1}$$

Hence we will call denote this probability distribution as observed probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ as can be easily computed from the observed statistics. This probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ will be the object of study in the device-independent framework. Different device-independent protocols will have different goals, but they follow the same scheme: perform useful (quantum) information tasks based solely on the analysis of the probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$.

Let us remind where the power of the device-independent framework comes from. Notice that the probability distribution does not contain explicitly any specific information about the particular physical systems inside the black box, be them atoms, photons, Higgs bosons or hamsters. Notice also that it does not depend on the precise experimental implementations inside the black box: if they were atoms, would them be ionized and in a trap or forming part of a molecule or ensemble? and if photons, would them be propagating freely inside the box, in a waveguide or in a fiber?. Since the probability distribution does

not carry *any* of this information, the conclusions we will derive from analizing it will be independent from any of these aspects. This makes device-independent protocols extremely robust against experimental imperfections, errors and security loopholes as we will later see.

The device-independent framework aims at making as few assumptions as possible so that the derived results are as general and meticulously specified as they can be. The assumptions of the framework, hence, must be few and well formulated. For the moment however, the described probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ seems too generic and structureless. In what follows we will see how to impose a minimal set of assumptions well motivated from very general physical considerations. This will make the problem nontrivial and interesting.

2.2.1. Assumptions of the framework

Let us now make two very important assumptions over the general device-independent framework previously described.

Freedom of choice

This assumption states that the choices of measurements of all parties $\mathbf{x}$ must be completely independent from the state of system being sent, the state of devices or any other possibly unknown degrees of freedom. A good way to formalize this notion in the language of special relativity was introduced in [CR12b] under the term *free randomness*. For a given fixed causal structure, a random variable X is said to be free if it is uncorrelated with everything except for its causal future, that is, from all variables lying outside its future light cone. If we collect all those variables and term them collectively e, we require that $P(\mathbf{x}|e) = P(\mathbf{x})$.

This assumption of freedom of choice is central to the majority of the results in device-independent quantum information protocols so far, with a few exceptions such as [KPB06, BG10, CR12b, GMDLT$^+$13]. In Chapter 4 we will however study how to relax this assumption to its breaking point and see if one can perform device-independent protocols under an almost complete relaxation of the said assumption. For the moment however, we will assume its validity.

Impossibility of instantenous signaling

This assumption is the same as the 'no signaling principle' introduced in the previous Section but adapted to the device-independent framework. It states that the choice of input of one party cannot influence the outcome given at another party's location.

This can be physically motivated using special relativity that states that there is maximum speed at which any signal can propagate. Events that are space-like separated are, in the language of special relativity, causally disconnected, ie. no signal could have travelled from one location to the other and directly influence the latter. In our case we will demand that the measurement processes of all parties, which are the variables (x_k, a_k) for each party k, define space-like separated events. We are specifically excluding the possibility of a signal carrying the outcome and input choice of one party to the measurement device of another hence potentially influencing its outcome. Mathematically this amounts to the following set of constraints that are known as *no-signaling constraints*:

$$\sum_{a_k} P_{\text{obs}}(a_1, \ldots, a_k, \ldots a_N | x_1, \ldots, x_k, \ldots x_N) \tag{2.1}$$

is independent of x_k for all k.

Both assumptions, namely no-signaling and free choice, are of course related since no-signaling is what fixes a causal structure. Without no-signaling, the free choice assumption in the form that we introduced does not make sense [CR12b].

2.2.2. Different theories permit different correlations

In all device-independent protocols there is an assumption which is vital: what is the background theory assumed to hold. Different conclusions can arise from the same observed correlations $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ depending on the background theory (or physical principles) that is assumed. This includes answers to questions such as how much randomness is present in the outcomes of an experiment or whether these are secret and useful for a cryptographic task.

In particular, assuming a particular background theory $\mathcal{T}$ implies that all the parties involved in the protocol are constrained by the said theory, including the eavesdropper or adversary. Thus, one says that quantum theory is assumed to hold, it means that the eavesdropper strategy involves both a quantum state possibly entangled with the other user's system and measurements on it . The information that the adversary may gain must follow the rules of quantum mechanics. If we say, on the contrary, that just no-signaling is assumed to hold, the eavesdropper is allowed stronger correlations with the user's systems as the no-signalling constraints are less stringent. This is the reason why assumming a bigger, less restricted theory implies that less randomness can be certified on the users side: they must take into account an eavesdropper that is more powerful and potentially more correlated with their systems.

In what follows we will describe several sets of correlations corresponding to different background theories. They will be used throughout the thesis at different points.

Classical correlations

These correlations arise from the assumption that both the physical systems being sent and the measurement processes taking place at each location are well described by a so-called local-hidden-variable-model (LHVM). Such models are essentially deterministic in the following way. The models assume that, for every party j, the outcome a_j for a measurement x_j would be completely determined if the *complete* description of the system was known. We will denote by λ this complete description, historically called *hidden variable* [Bel64]. Therefore, $P(a_j|\mathbf{x}_j, \lambda) = \delta^a_{f(x_j,\lambda)}$ where $f(x_j, \lambda)$ is a deterministic function that maps both input and the complete state of the system λ to an outcome a_j. Several rounds of the experiment may correspond to different states λ that could be drawn from a probability distribution $p(\lambda)$. Hence, the family of probability distributions that admit a LHVM model is

$$P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = \int p(\lambda) \prod_{j=1}^{N} \delta^{a_j}_{f_j(x_j,\lambda)} d\lambda \tag{2.1}$$

We will denote by $\mathcal{C}$ the set of $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ that admit such a model. Notice that the family of observed correlations one can engineer using physical systems fully governed by classical theories (such as classical mechanics, electrodynamics, special and general relativity) are all in $\mathcal{C}$. Making a connection to the previous Section, theories that from a generalized probabilistic theory (2.1) point of view are well characterized by CPT (2.1.6) produce correlations that admit a local hidden-variable model.

There is an alternative way to define LHVM from an assumption called 'locality' or 'factorizability' which however end up describing the same set [BCP+14]). We will stick to the present definition as it is best suited for our purposes.

Quantum correlations

Quantum correlations are those probability distributions that admit a quantum model behind. For a quantum model to exist, there should exist some quantum system ρ, which mathematically is a positive semidefinite operator $\rho \geq 0$ of unit trace $\text{tr}(\rho) = 1$ acting on the Hilbert space of N parties $\mathcal{H}_{A_1} \otimes \ldots \otimes \mathcal{H}_{A_N}$ and some measurement operators $O^{x_j}_{a_j}$ for every a_j and x_j which are also positive

semidefinite operators $O_{a_j}^{x_j} \geq 0$ each acting on the Hilbert space of the corresponding party and fulfiling $\sum_j O_{a_j}^{x_j} = \mathbb{I}$ where $\mathbb{I}$ is the identity matrix. These mathematical conditions ensure that probabilities are positive and normalized. The probability distribution of a quantum model follows the Born rule:

$$P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = \text{tr}(\rho \bigotimes_{j=1}^{N} O_{a_j}^{x_j}) \tag{2.1}$$

We will denote by $\mathcal{Q}$ the set of probability distribution that admit such a model.

Quantum correlations with dimension constraints

The set of quantum correlations with restricted dimension are those that, in analogy with the quantum set, admit a quantum model with a multipartite state $\rho \geq 0$ and measurement operators $O_{a_j}^{x_j} \geq 0$ where, however, the reduced state of some of the parties can be described by a state acting on a Hilbert space of dimension at most d. Different set of quantum correlations with restricted dimensions can be describe depending on which parties hold the dimension constraint and how strong it is. An example of a set of quantum correlations with dimension constraints would be

$$P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = \text{tr}(\rho \bigotimes_{j=1}^{N} O_{a_j}^{x_j}) \tag{2.1}$$

where ρ acts on a Hilbert space $\mathcal{H}_{A_1} \otimes \ldots \otimes \mathcal{H}_{A_N}$ with constraints such as $\dim(\mathcal{H}_{A_1}) \leq 3$ and $\dim(\mathcal{H}_{A_3}) \leq 2$. As we will explain in detail in Chapter 7, we will be interested in the convex hull of all such correlations. We will denote such sets by explicitly saying in what Hilbert space dimension the local states live in. For example, for the bipartite scenario where each party holds a qubit, we will denote the set by $Q(\mathbb{C}^2)$.

Locally quantum correlations

In analogy with Section 2.1.6 we will now describe the set of correlations that admit a locally quantum description. By locally quantum description we mean that, for each party $j \in \{1, \ldots N\}$, one can assign a valid quantum state $\rho_j \geq 0$ and measurement operators $O_{a_j}^{x_j} \geq 0$ fulfilling $\sum_{a_j} O_{a_j}^{x_j} = \mathbb{I}$ for all x_j. However, there need not exist neccessarily a *global* quantum state $\rho \geq 0$. Let us denote by W the global N-partite state of the system. We demand for the global state that all local measurements lead to valid probabilities. Thus, we must demand that $\text{tr}(W \otimes_{j=1}^{N} O_{a_j}^{x_j}) \geq 0$ for all $O_{a_j}^{x_j} \geq 0$. It is clear from this condition that all local

states must be valid quantum states ie. $\mathrm{tr}_{A_1 \ldots A_{j-1}, A_{j+1} \ldots A_N}(W) = \rho_j \geq 0 \ \forall j$. Probabilities that have a locally quantum model can be written as

$$P_{\mathrm{obs}}(\mathbf{a}|\mathbf{x}) = \mathrm{tr}(W \bigotimes_{j=1}^{N} O_{a_j}^{x_j}) \tag{2.1}$$

We will denote by $\mathcal{LQ}$ the set of probability distributions that admit such a model.

Maximally nonlocal correlations

The set of maximally nonlocal correlations is composed of all probability distributions that respect the no-signaling principle, that is, that fulfil the no-signaling constraints (2.2.1). It is the biggest set of correlations of all the sets described in this Section, as all of them respect the no-signaling principle and fulfil other additional constraints. In the following Section we will see why we call this set the *maximally nonlocal* set of correlations and we will denote it by $\mathcal{NS}$.

2.2.3. Bell nonlocality

Once the previous sets of correlations have been defined, we can define Bell nonlocality as simply the fact that $\mathcal{Q} \supset \mathcal{C}$. That is, that the set of quantum correlations is strictly larger than the set of classical correlations ie. there are quantum correlations that do not admit a LHV model. This simple statement hides in my opinion one of the most profund scientific discoveries in the history of physics. Moreover, along with its fundamental importance, nonlocality is fueling what has been called the *second quantum revolution* [DM03, Bel87]

To see why nonlocality is so remarkable, let us ponder the fact that quantum theory gives probabilistic predictions for measurement outcomes. All physical theories until quantum theory, ranging from classical mechanics to general relativity, were deterministic with no room for intrinsic unpredictability. Indeed, any source of uncertainty or unpredictability was always explained thanks to experimental errors, ignorance about the exact value of the parameters describing the experiment or a too strong approximation of the subset of degrees of freedom that are relevant to the experiment. Moreover, this ignorance could be arbitrarily reduced by using more precise equipment and improving the control all the relevant variables. Anyhow, the equations of evolution and the predictions for measurements were fully deterministic. This agrees with our own intuition about the world, were things exist and have definite values irrespective of whether we are observing them or not.

It was hence quite surprising that quantum theory was formulating its predictions in terms of probabilities solely. A very natural question at the time was whether quantum theory was really a *complete* theory. Einstein, Podolsky and Rosen [EPR35] thought that quantum theory would be superseeded in the future by a new theory that restored determinism making definite predictions. They had a famous scientific debate with Bohr [Boh35] who argued otherwise.

Nonlocality, discovered by John Bell in 1964 [Bel64] can be thought of as a negative answer to this question. Bell showed that quantum correlations are not compatible with the combination of the following three assumptions that we introduced: impossibility of instanteous signaling 2.2.1, freedom of choice 2.2.1 and the existence of some complete description of the state and measurement process λ that if known would allow for deterministic predictions (determinism) (2.2.2). The said complete description λ can be thought of as being written in the language of this new future theory that would presumably superseed quantum theory. The assumptions of no-signalling and free choice are commonly believed to be weaker assumptions than determinsm, which is usually the one to be dropped (with some notable exceptions as [Boh52]).

A major implication of the experimental observation of nonlocality (such as in [ADR82] modulo the detection loophole) is hence that, unless we drop one of the first two assumptions, *all future physical theories* will incorporate unpredictability as an intrinsic feature.

We now present a simple proof that $\mathcal{Q} \supset \mathcal{C}$ by stating Bell theorem in the flavour of one of the most famous witnesses of nonlocality: the Clasuer-Horne-Shimony-Holt (CHSH) Inequality [CHSH69].

Theorem 2.4 (Bell theorem via CHSH inequality)**.** *Let $P_{\text{obs}}(a,b|x,y)$ be an observed probability distribution in a device-independent scenario characterized by the triple $(2,2,2)$, where all variables take dichotomic values: $a,b \in \{+1,-1\}$ and $x,y \in \{0,1\}$. Let us denote by $\langle A_x B_y \rangle = \sum_{a,b} a \cdot b P_{\text{obs}}(a,b|x,y)$. Consider the following linear combination of elements in P_{obs}*

$$\mathrm{B}(P_{\text{obs}}) = \langle A_0 B_0 \rangle + \langle A_1 B_0 \rangle + \langle A_0 B_1 \rangle - \langle A_1 B_1 \rangle$$

For all $P_{\text{obs}} \in \mathcal{C}$ the following inequality holds $\mathrm{B}(P_{\text{obs}}) \leq 2$. However, there exist $P_{\text{obs}} \in \mathcal{Q}$ such that $\mathrm{B}(P_{\text{obs}}) = 2\sqrt{2} > 2$.

Proof. The bound of classical correlations $\mathcal{C}$ can be easily checked by hand since there are only 8 different extremal (and hence deterministic) probabilities according to equation (2.2.2). For quantum correlations it is enough to note that by measuring the maximally entangled state $|\psi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ with qubit measurements specified by the operators: $\hat{A}_x = O^x_{a=+1} - O^x_{a=-1} = \mathbf{a}_x \cdot \boldsymbol{\sigma}$ where

$\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices and $\mathbf{a}_x$ are unit vectors in $\mathbb{R}^3$. Similarly with Bob's measurements, parametrized analogously by the unit vectors $\mathbf{b}_y$. One can check that by measuring the singlet one obtains $\langle A_x B_y \rangle = -\mathbf{a}_x \cdot \mathbf{b}_y$. Hence, the inequality takes value $\mathrm{B}(P_{\mathrm{obs}}) = -\mathbf{a}_0 \cdot (\mathbf{b}_0 + \mathbf{b}_1) - \mathbf{a}_1 \cdot (\mathbf{b}_0 - \mathbf{b}_1)$. By choosing Bob's measurements to be given by $\mathbf{b}_y = -\frac{1}{\sqrt{2}}(\mathbf{a}_0 + (-1)^y \mathbf{a}_1)$ one finds $\mathrm{B}(P_{\mathrm{obs}}) = 2\sqrt{2}$. □

In what follows we will develop the mathematical theory behind the characterization of nonlocality. We will introduce the oft-used Bell inequalities and see that the CHSH inequality given in the previous Theorem is just one example.

2.2.4. Geometry of correlations and Bell inequalities

Observed probability distributions $P_{\mathrm{obs}}(\mathbf{a}|\mathbf{x})$ appearing in a (N, M, d) device-independent scenario can be thought of as vectors in Euclidean space $P_{\mathrm{obs}} \in \mathbb{R}^{(dm)^N}$ where every component corresponds to a specific combination of inputs and outputs $(\mathbf{a}|\mathbf{x})$. It is useful to study the geometrical form of the different sets of correlations described in 2.2.2 have and how to characterize them. First, let us write two simple mathematical conditions fulfiled by all sets

- Positivity:

$$P_{\mathrm{obs}}(a_1, \ldots, a_N | x_1, \ldots, x_N) \geq 0 \quad \forall (a_1, \ldots, a_N) \text{ and } (x_1, \ldots, x_N) \tag{2.1}$$

- Normalization:

$$\sum_{a_1, \ldots, a_N} P_{\mathrm{obs}}(a_1, \ldots, a_N | x_1, \ldots, x_N) = 1 \quad \forall x_1, \ldots x_N \tag{2.1}$$

As we shall see, with the exception of the quantum correlations with dimension constraints, all the sets of correlations described in 2.2.2 share another property: they are convex sets. We encountered convex sets in the previous Chapter on GPTs 2.1 where we saw that states $\mathcal{S}$, transformations $\mathcal{T}$ and measurements μ were also convex sets.

Convex sets can be defined as the *convex hull* of its extreme points or vertices. Among the different convex sets, those with a finite number of extreme points are of great importance for nonlocality. Mathematically, they are called *convex polytopes* and are a higher-dimensional generalization of convex polyhedra, such as a square or a triangle.

Definition 2.5 (Convex polytope)**.** *A convex polytope Ω can be defined by the following two equivalent representations called $\mathcal{H}$ (half spaces) and $\mathcal{V}$ (vertices):*

$\mathcal{V}$-representation: The convex hull of a finite set $X = \{P^1, \dots, P^n\}$ of points in $\mathbb{R}^d$:

$$\Omega = \text{conv}(X) := \left\{ \sum_i \lambda_i P^i \Big| \lambda_i \geq 0, \sum_i \lambda_{i=1}^n = 1 \right\}$$

$\mathcal{H}$-representation: A bounded solution set of a finite system of inequalities:

$$\Omega = \Omega(B, C) := \left\{ P \in \mathbb{R}^d \Big| B_i^T \cdot P \leq C_i \text{ for } 1 \leq i \leq m \right\}$$

where $B \in \mathbb{R}^{m \times d}$ is a real matrix with rows B_i^T, and $C \in \mathbb{R}^m$ is a real vector with entries C_i. $B_i^T \cdot P \leq C_i$ are called facets of the polytope. Boundness means that there exists a natural number N such that $||P|| \leq N$ holds for all $P \in \Omega$.

Classical correlations– Importantly, $\mathcal{C}$ is a convex polytope defined in the $\mathcal{V}$-representation in equation (2.2.2). The set of facets of the convex polytope $\mathcal{C}$ that would define its $\mathcal{H}$-representation is the set of tight *Bell inequalities*. As can be seen in Figure 2.2.4 a tight Bell Inequality separates neatly classical correlations from nonlocal correlations and is a standard tool to witness the presence of non-locality.

Quantum correlations– $\mathcal{Q}$ are convex bodies but not polytopes since they have an infinite amount of extreme points. A proof that quantum correlations are convex goes as follows. Imagine two N-partite probabilities P_1 and P_2 belong to $\mathcal{Q}$. That means that they admit a quantum model, ie. there exists two quantum states ρ_1, ρ_2 and sets of measurements $\{O_{a_j}^{x_j}\}_{(1)}$ and $\{O_{a_j}^{x_j}\}_{(2)}$ that generate P_1 and P_2 via the Born rule. We can generate any convex mixture of correlations $qP_1 + (1-q)P_2$ by simply preparing the quantum state $q\left|0\dots0\right\rangle\left\langle 0\dots0\right|_{\text{anc}} \otimes \rho_1 + (1-q)\left|1\dots1\right\rangle\left\langle 1\dots1\right|_{\text{anc}} \otimes \rho_2$ with the help of N ancillary flag qubits, each of which is sent along with the corresponding system. The measurement scheme for each device involves first measuring the flag qubit and then measuring the second system with the measurement operators $\{O_{a_j}^{x_j}\}_{(1)}$ or $\{O_{a_j}^{x_j}\}_{(2)}$ depending on the outcome of the flag qubit.

A proof that $\mathcal{Q} \supset \mathcal{C}$ can be derived from two facts. First, all classical correlations admit a quantum model. To construct one, we simply need to simulate the local deterministic probability $P(a|x) = \delta^a_{f(x,\lambda)}$ via measuring the quantum systems $|f(x,\lambda)\rangle$ in the orthonormal basis $\{|i\rangle\}, i = 0\dots, d-1$. Second, Bell theorem 2.4 proves that there are some $P_{\text{obs}} \in Q$ but $\notin \mathcal{C}$.

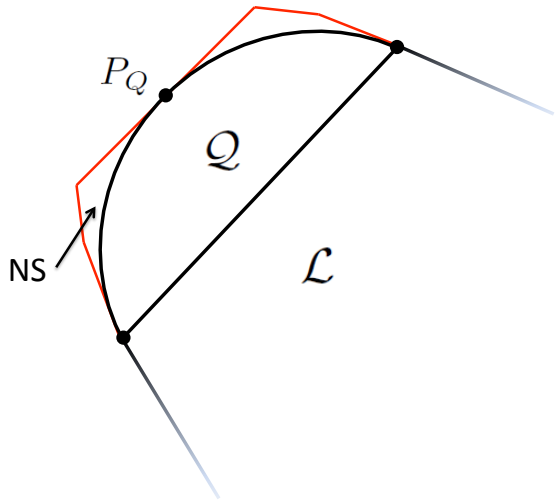

Figure 2.2.: Bell Inequality with quantum violation. In increasingly larger sizes, the local set is denoted by $\mathcal{L}$, the quantum set by $\mathcal{Q}$ and the no-signaling set by $\mathcal{NS}$. Note that each set described includes all the previous ones. The quantum set includes the local set and the no-signaling set includes the quantum set.

Quantum correlations with dimension constraints–The set of Q with dimension constraints is, on the contrary, not convex [PV09c]. Looking at the proof that quantum correlations are convex, we used the fact that we can prepare a state in an enlarged Hilbert space of system plus ancilla. If however, the dimension of the state has an upper bound, this proof does not hold. It is hence a possibility that the set of correlations ceases to be convex, and this is indeed what happens [PV09c]. We will later see in Chapter 7 however, that the convex hull of such a set is what is relevant. To see why this is the case, consider that, if quantum systems of bounded dimension are the resource, the users of a Bell test can receive probabilistic mixtures of such states. We can hence think of these sets as convex subsets of the set $\mathcal{Q}$ which in general also allow for violations of Bell Inequalities.

Locally quantum correlations– Let us now address the set of locally quantum correlations $\mathcal{LQ}$. A surprising result in [AAC$^+$10, BBB$^+$10] proved that although the set of locally quantum states W that appear in the Born rule of equation (2.2.2) is bigger than the set of quantum states (remember that W does not need to be fully positive) the set $\mathcal{LQ}$ of correlations is equal to the quantum

set $\mathcal{LQ} = \mathcal{Q}$ for all *bipartite* correlations. This has the important implication that there is no way to distinguish both theories in a device-independent way by any pair of parties. Note however that for three or more parties this ceases to be the case. In particular, there are correlations that are in the locally quantum set $P_{\text{obs}} \in \mathcal{LQ}$ but not in the quantum $P_{\text{obs}}\mathcal{Q}$ [AAC+10] which proves that in general $\mathcal{LQ} \supset \mathcal{Q}$. The set of correlations $\mathcal{LQ}$ is also convex. This can be proved in a similar fashion to the previous proof of convexity of quantum correlations.

Maximally nonlocal correlations– Finally we address the case of maximally nonlocal correlations $\mathcal{NS}$. The set is also a convex polytope defined in $\mathcal{H}$-representation by the inequalities that correspond to positivity (2.2.4) as well as the equality constraints that correspond to the no-signaling constraints (2.2.1) and normalization (2.2.4). These conditions restrict $P_{\text{obs}} \in \mathcal{NS}$ to live in a *hyperplane* of $\mathbb{R}^{(Md)^N}$, hence decreasing its dimension.

The relation of $\mathcal{NS}$ with the quantum and classical set is $\mathcal{C} \subset \mathcal{Q} \subset \mathcal{NS}$. In order to prove it, since we have already proved that $\mathcal{C} \subset \mathcal{Q}$, we need to show that $\mathcal{Q} \subset \mathcal{NS}$. The proof of inclusion is simply to notice that quantum correlations fulfill the no-signaling constraints (2.2.1)

$$\sum_{a_k} \text{tr}(\rho O_{a_1}^{x_1} \otimes \ldots \otimes O_{a_k}^{x_k} \otimes \ldots O_{a_N}^{x_N}) = \tag{2.2}$$

$$\text{tr}(\rho O_{a_1}^{x_1} \otimes \ldots \otimes (\sum_{a_k} O_{a_k}^{x_k}) \otimes \ldots O_{a_N}^{x_N}) = \tag{2.3}$$

$$\text{tr}(\rho O_{a_1}^{x_1} \otimes \ldots \otimes (\mathbb{I}) \otimes \ldots O_{a_N}^{x_N}) \tag{2.4}$$

which is independent of x_k. Where for the first equality we have used the linearity of the trace and for the second that the set of POVMS $\{O_{a_k}^{x_k}\}$ are a partition of the identiy, ie. $\sum_{a_k} O_{a_k}^{x_k} = \mathbb{I}$ as explained in 2.2.2. For the proof of strict inclusion, it is enough to see that there are correlations beloging to $\mathcal{NS}$ which do not belong to $\mathcal{Q}$. The most famous example is the so-called Popescu-Rohrlich (PR) box [PR94], defined by the probability distribution $P_{PR}(a, b|x, y)$ which is equal to $1/2$ whenever $a \oplus b = x \cdot y$ and 0 otherwise and the addition is taken modulo two. One can easily check that this probability distribution achieves a Bell violation of 4 in the CHSH inequality (2.4). Tsirelson proved that the maximum quantum violation for the CHSH inequality is $2\sqrt{2}$ and that it can be obtained with qubits [Tsi80], as we showed previously as well.

As a remark, determining whether some correlations P_{obs} belong to a particular set can be casted as a *feasibility problem*. If the set is a convex polytope, like both the classical and the maximally nonlocal sets of correlations, the problem becomes a *linear problem* and can be solved via linear programming. In the

case of quantum correlations and locally quantum correlations, the feasibility problem can be casted as hierarchy of semidefinite programs (see [NPA07] and Chapter 7 respectively).

2.3. Randomness from a physicist perspective

As previously explained, quantum theory and unpredictability have been linked from the early days of the theory. Debate about the completeness of quantum mechanical predictions and the proof by John Bell that quantum correlations did not admit a deterministic completion already pointed out at this deep connection between nonlocality and 'true' randomness.

This connection was however taken to a more quantitative level in [Col07] and even more so in [PAM+10a] where an explicit protocol that produces a large amount of random bits starting with only a small amount of them was developed. The protocol measures entangled quantum particles to generate correlations that violate a Bell Inequality.

In what follows we will state the basics of randomness certification from purely operational means. The setting is the standard one for device-independent protocols: A Bell test is performed repeatedly among N parties and the resulting statistics is given by a probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$, where $\mathbf{a} = (a_1, \ldots, a_N)$ and $\mathbf{x} = (x_1, \ldots, x_N)$ are the string of outcomes and measurement inputs of the parties involved.

In Section 2.2.2 we saw that a vital assumption was the set of correlations assumed to be physical ie. the quantum set $\mathcal{Q}$, the set of all correlations respecting the no-signaling principle $\mathcal{NS}$ or some other set of correlations. We will denote by T the set of correlations we assume. Results about device-independent randomness certification are hence theory-dependent, but not model-dependent. That is because they assume the theory governing the physics of the experiment but not the specific model at work. A model would involve stating the precise mathematical implementation of an experiment ie. the state produce, the action of the beamsplitters and measurements devices, etc.

To begin with we will distinguish two different notions of randomness by relating them to the predictability of two different experimental situations.

2.3.1. Observed randomness

Imagine a provider gives you a multipartite black box from which you know nothing. By probing the box with different inputs $\mathbf{x}$ and collecting the outputs $\mathbf{a}$ you find out that it behaves according to the probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$. We would like to capture how predictable those outcomes are by asking: what

is the most likely outcome when your input is $\mathbf{x}$? That is what we will define now:

Definition 2.6 (Observed randomness)**.** *The observed randomness G_{obs} on $\boldsymbol{a}$ for measurements $\boldsymbol{x}$ is the randomness computed directly from the observed statistics P_{obs}. Operationally, this may be defined as the optimal probability of guessing the outcomes $\boldsymbol{a}$ for input $\boldsymbol{x}$,*

$$G_{\text{obs}}(\boldsymbol{x}, P_{\text{obs}}) = \max_{\boldsymbol{a}} \; P_{\text{obs}}(\boldsymbol{a}|\boldsymbol{x}). \tag{2.4}$$

We can likewise speak in terms of another measure of randomness, the min-entropy, which depends on the guessing probability through,

$$H_{\infty}^{\text{obs}}(\mathbf{x}, P_{\text{obs}}) = -\log_2 G_{\text{obs}}(\mathbf{x}, P_{\text{obs}}). \tag{2.4}$$

2.3.2. Intrinsic randomness

Let us now move to a definition of *intrinsic randomness*. Imagine the same situation as before, where an untrusted provider (or eavesdropper) gives you a multipartite black box whose behaviour is captured by $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$. However, you are now asked to perform a different task. Your goal is now to describe the predictability of the outcomes for the provider/eavesdropper of the black box.

Notice that even if the observed randomness is high, there might be some further parameters that we will collectively call e that, if known to the provider, could help him to better predict the outcomes of the experiment. Another way to look at it is that the observed correlations P_{obs} are in fact a preparation in terms of 'maximum knowledge' probability distributions within theory T. Since theory T must be convex (we can always prepare mixtures by flipping a coin), these must be the extremal points P_e^{ex} of the set T. In our context, a particular preparation reads

$$P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = \sum_e p(e) P_e^{\text{ex}}(\mathbf{a}|\mathbf{x}) \tag{2.4}$$

where the P_e^{ex} are extremal points of the set T of correlations.

Since we do not know which is the actual preparation that the provider used to produce the observed statistics, we should consider all possible preparations of the statistics and define the intrinsic randomness of outcomes $\mathbf{a}$ by the *worst case*, optimizing over all possible preparations of P_{obs} so as to minimize the randomness on $\mathbf{a}$.

Definition 2.7 (Intrinsic randomness)**.** *The intrinsic randomness G_{int}^T on $\boldsymbol{a}$ for input measurements $\boldsymbol{x}$ is the randomness seen by a party (the provider) ruled*

by a theory T and in possession of all the information about the preparation $\{p(e), P_e^{\text{ex}}\}$ of some observed correlations P_{obs} in the worst-case scenario

$$G_{\text{int}}^{T}(\boldsymbol{x}, P_{\text{obs}}) = \max_{p(e), P_e^{\text{ex}} \in T} \sum_e p(e) G_{\text{obs}}(g, \boldsymbol{x}, P_e^{\text{ex}})$$

subject to:

$$\sum_e p(e) P_e^{\text{ex}}(\boldsymbol{a}|\boldsymbol{x}) = P_{\text{obs}}(\boldsymbol{a}|\boldsymbol{x})$$

where $G_{\text{obs}}(\boldsymbol{x}, P_e^{\text{ex}}) = \max_{\boldsymbol{a}} \; P_e^{\text{ex}}(\boldsymbol{a}|\boldsymbol{x})$ is also the intrinsic randomness of P_e^{ex}, since intrinsic and observed randomness must coincide for extremal points

Notice that for classical correlations $P_{\text{obs}} \in \mathcal{C}$, even if any value of G_{obs} can be obtained, the intrinsic predictability is always maximal since, for the extremal points of the set $G_{\text{obs}}(\mathbf{x}, P_e^{\text{ex}}) = \max_{\mathbf{a}} \; \delta^{\mathbf{a}}_{f_\lambda(\mathbf{x})} = 1$.

Another interesting remark is that $G_{\text{int}} \geq G_{\text{obs}}$ since a possible preparation of the correlations by the eavesdropper is of course the trivial one, sending P_{obs} directly.

In the same fashion we can also express the intrinsic randomness in terms of the min-entropy by

$$H_{\infty}^{\text{int}T}(\mathbf{x}, P_{\text{int}}) = -\log_2 G_{\text{int}} T(\mathbf{x}, P_{\text{obs}}). \tag{2.2}$$

2.4. Randomness from a computer scientist perspective

In the previous Section we have gone through the basic principles of randomness certification from a device-independent perspective. Importantly, all the previous analysis had in mind a notion of randomness that is natural in physics: a process such as a coin toss, a position measurement of a chaotic system or a measurement of spin, contains intrinsic randomness if the best model for any observer is given by a probability distribution with support in more than one outcome. For instance, the strongest notion of randomness for a physicist is a balanced probability distribution among the outcomes, ie. $p(\text{heads}) = p(\text{tails}) = 1/2$.

On the other hand, randomness has a quite different meaning in computer science. For a computer scientist, randomness is not a property of the potential outcomes of a process best captured by a probability distribution, but a syntatic propery of a particular *infinite sequence* X. To help to illustrate better the concept consider the following three sequences

00000000000000000000000

01010101010101010101010

01100101110110110001010

Although all of them are 23 bits long and have equal chance of coming out of a fair coin, probability $p = 2^{-23}$, there is a difference in how to *describe* these sequences. The first sequence, for instance, can be described by saying '23 times zeros'. The second sequence is slightly more complex, but still rather simple. We can describe it by saying if $x = 0$ if x is an odd postion, and $x = 1$ otherwise. The third sequence however, does not seem so simple to describe in fewer words than stating the sequence itself. Roughly speaking, we will see that an i*nfinite* sequence is truly random in the algorithmic sense if the shorter description of it is itself, ie. it is not 'compressible'. In this Section we will elaborate more in this notion.

In Chapter 8 will use the concepts and tools that we will now introduce in order to study how both types of randomness interact in the field of quantum physics and what are the implications. The following introduction follows closely the exposition given at [FS14]. The reader is refered to [LV09] for a more in depth discussion of the topics.

2.4.1. Introduction to computability theory and Turing machines

Computability theory studies which functions can be calculated by algorithmic means and which cannot, and, more generally, their degree of uncomputability. The field emerged in the 1930s with the independent works of Alan Turing [Tur36] and Alonzo Church [Chu36] who introduced two equivalent formalizations of the intuitive concept of *algorithm*. The fact that both models of computation turned out to be equivalent lead Stephen Kleene [Kle67] to postulate what is now known as the Church-Turing thesis:

> *any function 'naturally to be regarded as computable' (i.e. calculable by algorithmic means) is computable by the formal model of Turing machines.*

This thesis has been greatly strengthened by the fact that all the formal models of computation defined so far have been shown to be at most as expressive as the classical Turing machines in terms of the class of functions they compute.

Turing machines are an abstraction of a computer. Indeed, they are like our desktop machines but with an infinite amount of memory. A Turing machine works as follows. It has two parts, a semi-infinite tape partitioned in infinitely many finite cells and a writing head composed of a writing mechanism and an internal finite memory in which two things are saved: first, a finite list of 'transition rules' or *program* and second, an internal state out of a finite set, say $\{A, B, C, \ldots, \text{Halt}\}$.

The finite *input* $x \in \mathbb{N}$ to the Turing machine is written in the first cells of the tape using a finite alphabet, say binary $\{0, 1\}$. The Turing machine initialized with its internal state in A will work as follows: Start in the first cell and read its value, say 0. Search in its internal memory for a transition rule saying something like $\delta(A, 0) = (B, 1, \rightarrow)$ which means 'if the cell reads 0 and the internal state is A, change your internal state to B, rewrite the cell with the value 1 and move one step to the right. The machine will work in this fashion until a transition rule like this $\delta(B, 1) = (\text{Halt}, 1, \rightarrow)$ is implemented. When the internal state is in Halt the machine stops. The sequence written in the tape will the the output of the computation.

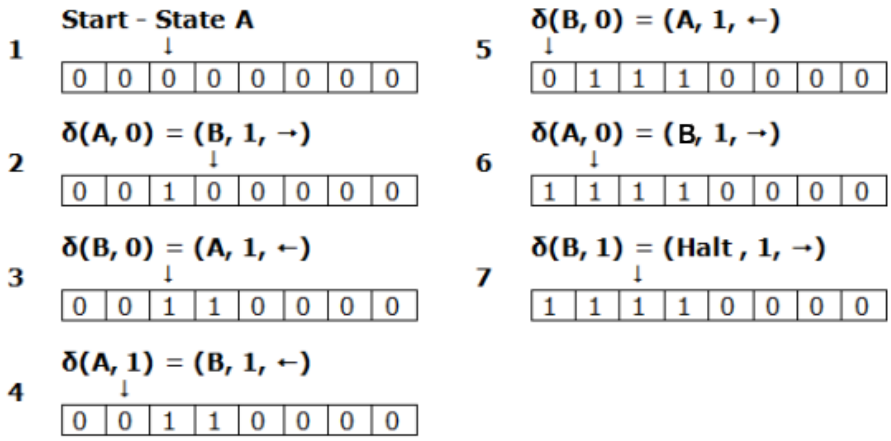

Figure 2.3.: Turing machine evolution from Start to Halt. The machine works as follows. Starts with its internal state in A. The head reads the current symbol and looks fro the transition rule that matches both the symbol and the internal state. Then performs what the transition rule dictates: changes its internal state, rewrites the position and moves either to right or left. Afterwards the process starts again until the internal state changes to Halt.

Computability theory defines everything that can be computed using a Turing machine as *computable*. Importantly, because all the components of the Turing machine are *finite*, one can always find a umambiguous enconding of all its elements so as to enumerate all Turing machines TM by a natural number $\sharp \mathrm{m} \in \mathbb{N}$. For instance, TM $\sharp m$ represents the m-th Turing machine.

One could now ask the question: are all functions $f : \mathbb{N} \to \{0, 1\}$ computable? The basis for computability theory is the negative answer that Turing found to this question. The proof is a gem of mathematical theory both for its simplicity and its far reaching consequences. It works by noticing that there are some programs implementing some functions $f(x), x \in \mathbb{N}$ that do not halt on some inputs x, but in general one cannot know which ones will halt *a priori*: in other words, the halting problem is not computable.

The proof of the uncomputability of the halting problem or correspondingly the non-existence of a Turing machine solving the Halting problem goes by contradiction. Suppose that the halting problem was computable, which means there exists a Turing machine H computing the following function $\mathrm{Halt}(x)$

$$\mathrm{Halt}(x) = \begin{cases} \text{True} & \text{if TM } \sharp\text{x halts on input x} \\ \text{False} & \text{if TM } \sharp\text{x does not halt on input x} \end{cases}$$

Then, it should exist also a Turing machine implementing the following function $f(x)$ given in pseudocode

```
If Halt(x) = True then
loop forever
Elseif Halts(x) = False then
halt
endif
```

But then, $f(\sharp f)$ halts $\Leftrightarrow$ $f(\sharp f)$ does not halt. In words, by giving as an input the program that computes $f(x)$ encoded in the natural number $\sharp f$ to the Turing machine computing $f(x)$ itself, one can create a contradiction: if the program halts it does not halt and viceversa. Hence the initial assumption must be incorrect and $\mathrm{Halt}(x)$ is not computable $\square$.

Therefore not all functions are computable. Computability actually cares about *partial* functions $f : \mathbb{N}^k \to \mathbb{N}$. Partial means that they can be undefined in some of the inputs. The notation we will use will be as follows: $f(x) \downarrow$ means $f(x)$ is defined, and hence a natural number. Likewise, $f(x) \uparrow$ means $f(x)$ is undefined. We will denote by $\mathrm{dom}(f) = \{x | f(x) \downarrow\}$. If $\mathrm{dom(f)} = \mathbb{N}^k$ we say that the function is total. A function is called Turing-computable if there exists a Turing machine that, reading input x, halts with $f(x)$ written as an output in the tape whenever $f(x) \downarrow$ and otherwise never halts.

2.4.2. Computability theory and infinite sequences

Now we will expand the introduction of computability to incorporate infinite sequences. The set of finite strings over the alphabet $\{0,1\}$ is denoted 2^* and ϵ denotes the empty string. The set of infinite sequences over $\{0,1\}$ is denoted 2^ω. If $S \in 2^\omega$ then $S \upharpoonright n$ denotes the string in 2^* formed by the first n symbols of S. If $x, y \in 2^*$ then $x \preceq y$ represents that x is a prefix of y. Any natural number n can be seen as a string in 2^* via its binary representation.

Let us first introduce formally the computing model used throughout this thesis. All in all, it is nothing but a particular model equivalent to a Turing machine we have just explained, thus having the same computational power as any computer with unbounded memory. Specifically, we consider Turing machines M with a reading, a working and an output tape (the last two being initially blank). The output of M for input $x \in 2^*$ is denoted $M(x) \in 2^*$, and if $t \geq 0$, $M_t(x) \in 2^*$ consists of the content of the output tape in the execution of M for input x by step t —notice that this execution needs not be terminal, that is, $M(x)$ needs not be in a halting state at stage t. A *monotone* Turing machine (see e.g. [DH10, S2.15]) is a Turing machine whose output tape is one-way and write-only, meaning that it can append new bits to the output but it cannot erase previously written ones. Hence if M is a monotone machine $M_t(x) \preceq M_s(x)$ for $t \leq s$. The computing model of monotone machines is equivalent to ordinary Turing machines, and for ease of presentation we work with the former.

A sequence $S \in 2^\omega$ is *computable* if there is a (monotone) Turing machine M such that for all n, $M(n) = S \upharpoonright n$. Equivalently, S is computable if there is a monotone machine M such that $M(\epsilon)$ "outputs" S, in the sense that

$$(\forall n)(\exists t)\ M_t(\epsilon) = S \upharpoonright n. \tag{2.2}$$

Let $(M_i)_{i\geq 0}$ be an enumeration of all monotone Turing machines and let U be a monotone Turing machine defined by $U(\langle i, x\rangle) = M_i(x)$, where $\langle \cdot, \cdot \rangle : \mathbb{N}^2 \to \mathbb{N}$ is any computable pairing function (i.e. one that codifies two numbers in $\mathbb{N}$ into one and such that both the coding and the decoding functions are computable). The machine U is universal for the class of all monotone machines. In other words, U is an interpreter for the class of all monotone Turing machines, and the argument p in $U(p)$ is said to be a *program* for U, encoding a monotone Turing machine and an input for it.

The notion of computability makes sense when applied to infinite sequences, as any finite string can be trivially computed by a very simple (monotone) Turing machine which just hard-codes the value of the string. Any finite binary string can be extended with infinitely many symbols in order to obtain either

a computable or an uncomputable sequence. For instance, if s is a finite string then s followed by a sequence of zeroes is computable; however s followed by the (binary representation of) the halting problem [Tur36] is not. Since in finite time a Turing machine can only process finitely many symbols, one cannot decide in finite time if an infinite sequence is computable or not.

2.4.3. Algorithmic complexity and randomness

There are several notions of algorithmic randomness but one of the most standard ones is the so called 1-randomness or Martin-Löf randomness. There are mainly three equivalent definitions of $1-$randomnes according to three different intuitive notions of what a random infinite sequence is. An infinite sequence is random if it passes *all* performable statistical tests [Mar66] or if there does not exist any effective strategy or martingale to win an unbounded amount of money betting on that sequence [Sch71] or, to finish with, if the sequence is essentially incompressible [Sch72, Lev73]. We will use the definition in terms of compressibility or Kolmogorov complexity of a sequence or prefix, since is the one of more relevance to us.

The Kolmogorov complexity of a sequence is, intuitively speaking, the lenght of the shortest program (or description) that can generate that sequence. The only problem with this intuitive definition is the question of what kind of programming language are we using. To convey a description, does it matter if we use Java, matlab, or English? To take into account this ambiguity we need to define a unique way of talking about the Kolmogorov complexity. That can be done refering to a Universal Turing Machine defined in the previous Section.

Plain Kolmogorov complexity

A Turing machine supplemented with an input describes a sequence. We say a machine N describes a sequence x if $N(p) = x$. We say that program p is the description of x for N.

Definition 2.8 (Kolmogorov complexity)**.** *Given a Turing Machine N, the plain Kolmogorov complexity relative to N, $C_N(x)$ is the lenght of the shortest program which describes x in N:*

$$C : 2^* \to \mathbb{N} \cup \{\infty\}$$

$$C_N(x) = \begin{cases} \min\{|p| : N(p) = x\} & \text{in case } \{|p| : N(p) = x \neq \emptyset\} \\ \infty & \text{otherwise} \end{cases}$$

We say that *the* plain Kolmogorov complexity is $C = C_U$ where U is the universal Turing machine. Every finite string w can be described by a program and therefore $C(w) \leq \infty$. It can be proved [LV08] that the plain Kolmogorov complexity is minimal in the sense that for every machine N we have that $\exists c : C(x) \leq C_N(x) + c \ \forall x$. Although the plain Kolmogorov complexity of a sequence is not computable, for finite sequences it can be approximated from above by running programs of smaller or equal lenght of the sequence, for longer and longer steps of time. The fact that is not computable can be related directly with the halting problem, as we do not know if waiting more some short program could output the desired finite sequence.

However, the plain Kolmogorov complexity has two main drawbacks. First, there are no infinite binary sequences X such that

$$(\exists d)(\forall x \prec X)C(x) > |x| + d \tag{2.2}$$

which would be a desirable property of incompressibility (and hence randomness) and second, it is not true that $C(\langle x, y \rangle) \leq C(x) + C(y) + d$ where d is a constant. One would like a definition of complexity to verify that the complexity of codifying the two strings together should be smaller or equal that the sum of the independent complexities, as there coud be correlations between both sequences that would allow to describe them jointly in a more succint form. The fact that these two conditions do not hold make the plain Kolmogorov complexity unsuitable for the purpose of defining $1-$randomness. We will see how to solve these two caveats with a different definition of complexity.

Prefix-free or self-delimiting machines

A set of words is *prefix-free* if in the case $w \in A$ then no proper extension of w is in A, ie:

$$w \in A \rightarrow (\forall v \neq \emptyset) wv \notin A$$

Definition 2.9 (Kolmogorov complexity)**.** *A machine M is prefix-free if its domain,* $\mathrm{dom}(M)$*, is prefix-free*

That is, in case $M(w)\downarrow$ is defined and $w \prec v$ then $M(v)\uparrow$ is not defined.

A way to look at a prefix-free machine, also called self-delimiting, is to imagine a Turing machine with two tapes. The finite input tape to read only and the working/output tape. Now the head starts in the leftmost cell of the input tape and, to halt it needs to be in the rightmost cell of the input tape. There are no blank or end marker. In that way, the machine is forced to accept strings

without knowing whether there are any more bits written in the input tape. The lenght should be somehow codified in the very same input. Attempt to go to the right of the rightmost input cell implies automatically the undefinition of the machine.

We can analogously define an universal prefix-free Turing machine as follows. Fix an effective enumeration of all prefix-free Turing machines $M_0, M_1, M_2, \ldots$. Let us fix a particular universal Turing machine $U(\langle e, x\rangle) = M_e(x)$. We are now ready to define the Prefix Kolmogorov complexity.

Definition 2.10 (Prefix Kolmogorov complexity)**.** *Given a prefix Turing Machine M, the prefix Kolmogorov complexity relative to M, $K_M(x)$ is the lenght of the shortest program which describes x in N:*

$$K : 2^* \to \mathbb{N} \cup \{\infty\}$$

$$K_M(x) = \begin{cases} \min\{|p| : M(p) = x\} & \text{in case } \{|p| : M(p) = x \neq \emptyset\} \\ \infty & \text{otherwise} \end{cases}$$

We say that *the* prefix Kolmogorov complexity is $K = K_U$ where U is the universal Turing machine fixed before. Again, every finite string w is describable by a program so $K(w) < \infty$.

Equipped with this notion of complexity we can finally define $1-$randomness

Definition 2.11. *An infinite sequence X is $1-$random if if (almost) every prefix is incompressible, that is*

$$(\exists c)(\forall n)\ K(X \upharpoonright n) \geq n - c \tag{2.2}$$

This notion of randomness, formulated here in terms of imcompressibility, will be the underlying notion used during this thesis.

3. Locally quantum theories

In the past decades, Quantum Information Theory has provided information tasks where quantum theory offers a fundamental advantage. However, the abstract and mathematical nature of its axioms prevents us from intuitively and physically understanding where the power of quantum theory comes from.

To gain a better understanding of quantum theory, it is helpful to compare and contrast it with other conceivable physical theories within the general probabilistic framework of GPTs 2.1. In this Chapter we will be interested in those theories which have the same local structure as quantum theory, but a potentially different global structure as defined in Chapter 2, Section 2.1.6. What singles out quantum theory from the class of *locally quantum* theories?

Here we will show that asking for *dynamical reversibility* is enough to single ot quantum theory from this space of theories and discuss the implications of this result for quantum computation. The results of this Chapter are based on the article [dlTMSM12].

3.1. Introduction

In a recent article, Barnum et al [BBB$^+$10] studied the correlations that could arise from locally quantum correlations. As we explained in the preliminaries Chapter 2.1.6, such theories may in general contain non-quantum states, corresponding to entanglement witnesses [HHH96] and one could then expect this set to be larger than the quantum set. They showed however that for two parties this was not case. Indeed, they showed that the set of bipartite correlations attainable in any locally quantum theory cannot be larger than the set of quantum correlations. Nonetheless, when three or more parties are considered, locally quantum theories were later shown to yield stronger than quantum non-local correlations [AAC$^+$10].

An important property of quantum theory which is not shared by all theories is the existence of a continuous reversible transformation between any two pure states. Demanding that the fundamental dynamics is reversible and continuous in time seems like a natural physical requirement, and is a key component of several recent axiomatic reconstructions of quantum theory [Har01, CDP11,

MM11, DB11]. Furthermore, an alternative theory yielding much stronger non-local correlations than quantum theory, known as no-signaling theories was recently shown to contain no reversible interactions [GMCD10]. Here we explore how reversibility constrains the global structure of quantum theory.

Locally quantum theories in the GPT formalism

Recall from 2.1 that in the GPT framework the state of a system is characterised by the outcome probabilities for some set of measurements, called *fiducial* measurements. This set is generally non-unique, but must be sufficient to derive the outcome probabilities for any other measurement. For qubits, we will take the fiducial measurements to be the three Pauli spin operators, which we denote by $\sigma_1, \sigma_2, \sigma_3$.

Joint systems are assumed to be completely characterised by the probability distributions for every combination of fiducial measurements on the component systems. This property was called *local tomography* in the preliminaries Chapter (Definition 2.2). The state of n qubits can therefore be represented by the conditional probability distribution $P(a_1, \ldots, a_n | x_1, \ldots, x_n)$ which gives the joint probability of obtaining results $a_1, \ldots, a_n \in \{+1, -1\}$ when measuring $x_1, \ldots, x_n \in \{\sigma_1, \sigma_2, \sigma_3\}$ respectively on the n systems. In order for the probabilities of outcomes on a single system to be well-defined, and to prevent instantaneous signalling between parties, it is important that they fulfil the *no-signaling constraints* stated in equation (2.2.1).

It is shown in [AAC$^+$10] that any state satisfying the no-signaling constraints can be represented in an analogous way to a standard quantum state, as a trace-1 Hermitian operator ρ on a 2^n-dimensional Hilbert space. The outcome probabilities for the fiducial measurements in turn are obtained in precisely the same way as in standard quantum theory, ie. through the Born rule $p(m) = \mathrm{tr}(\rho M_m)$. However, the operator ρ need not be positive in general. As we are considering local quantum theories, the set of separable states and the set of local measurements must be the same as in quantum theory. However, the entangled states and measurements may differ.

Key constraint: continuous and reversible dynamics

In general, the dynamics of a system can be controlled by adjusting its environment. Since we assume that the fundamental dynamics is reversible and continuous in time, the transformations implemented in this way form a connected group. In order to respect probabilistic mixtures, transformations must

act linearly on the state ρ as argued in 2.1. Note that we adopt an operational approach, so the allowed transformations include possibilities in which an ancilla is prepared, evolves jointly with the system, and is then discarded (in quantum theory the allowed transformations are the completely-positive trace-preserving (CPTP) maps).

The group of transformations on n qubits $\mathcal{G}$ must then obey two important conditions:

1. **Quantum theory holds locally:** The first condition is that it should contain the local unitary transformations $\mathcal{G}_{\text{loc}}$ for the qubits. A unitary transformation $U \in \text{SU}(d)$ acts on a d-level quantum system via the adjoint action $\text{ad}_U[\rho] = U\rho U^\dagger$. Hence for n qubits

$$\mathcal{G}_{\text{loc}} = \{\text{ad}_{U_1} \otimes \cdots \otimes \text{ad}_{U_n} : U_r \in \text{SU}(2)\} \subseteq \mathcal{G}.$$

2. **Consistency constraint:** The second condition on the group $\mathcal{G}$ is that the combination of preparing a product state, transforming it using any $G \in \mathcal{G}$, then performing a product measurement must yield a valid outcome probability, ie. between 0 and 1.

3.2. Results

The two conditions stated before allow us to prove our main technical result. The proof of our main result will be given later in (3.1). Informally, it states that there are only two possible locally quantum theories with non-trivial continously reversible dynamics. One is the usual quantum mechanics for n qubits, and the other is quantum mechanics with partial transposition. Both theories are equivalent in their computational power. We will later see how to eliminate the second possibility by a simple requirement: that all the qubits are identical type of systems

3.2.1. Two possible reversible and locally quantum theories

Theorem 3.1 (Two possible compatible theories)**.** *Let $\mathcal{G}$ be a connected group which acts linearly on the set of $2^n \times 2^n$ Hermitian matrices and satisfies:*

1. $\mathcal{G}_{\text{loc}} \subseteq \mathcal{G}$,
2. $\text{tr}\Big((\mu_1 \otimes \cdots \otimes \mu_n)\, G[\rho_1 \otimes \cdots \otimes \rho_n]\Big) \in [0,1]$ *for any $G \in \mathcal{G}$ and any qubit states μ_r and ρ_r.*

If $\mathcal{G} \neq \mathcal{G}_{\text{loc}}$ then there exist: $G \in \mathcal{G}$, a re-ordering of the qubits, an entangling unitary $U \in \text{SU}(4)$, and a single-qubit state σ, such that one of the following possibilities holds:

1. $G[\rho_{12} \otimes \sigma^{\otimes(n-2)}] = \text{ad}_U[\rho_{12}] \otimes \sigma^{\otimes(n-2)}$,
2. $G[\rho_{12} \otimes \sigma^{\otimes(n-2)}] = (T_1 \circ \text{ad}_U \circ T_1)[\rho_{12}] \otimes \sigma^{\otimes(n-2)}$,

for any Hermitian matrix ρ_{12}, where T_1 is the partial transposition operation on qubit 1.

This theorem means that in any locally quantum theory of n qubits which allows for a (not necessarily quantum) interaction, $\mathcal{G} \neq \mathcal{G}_{\text{loc}}$, there exist a particular pair of qubits, i and j, on which we can implement either the quantum transformation $G_{ij}^U = \text{ad}_U$, or the non-quantum $H_{ij}^U = (T_i \circ \text{ad}_U \circ T_i)$. As any bipartite entangling unitary plus local unitaries are sufficient to generate all unitary transformations [Har09b], when we can implement G_{ij}^U we can implement all unitary transformations on qubits i and j. If we can implement H_{ij}^U instead, we note that for any $V_i \in SU(2)$, there exists $V_i' \in SU(2)$ such that $\text{ad}_{V_i'} = (T_i \circ \text{ad}_{V_i} \circ T_i)$. Hence by sequences of local transformations and H_{ij}^U operations we can implement $(T_i \circ \text{ad}_V \circ T_i)$ on qubits i and j, for any $V \in SU(4)$.

Now consider the additional assumption that all qubits are identical, in the sense that they are the same type of system (although they may have different states). This means that given $(n + m)$ qubits (for any $m \geq 0$), for any $G \in \mathcal{G}$, and any permutation π of the qubits, $\pi \circ (G \otimes I^{\otimes n}) \circ \pi^{-1}$ is an allowed transformation. In this case we can prove a stronger result.

3.2.2. Only quantum theory survives

Main result: Consider any locally-tomographic theory in which the individual systems are identical qubits. If the theory admits any continuous reversible interaction between systems, then the allowed states, measurements, and transformations must be identical to those in quantum theory. This is the most important result of the Chapter. It shows the intimate relationship between the local and nonlocal structure of quantum theory through a physical constraint on the dynamics: that it must be continuously reversible.

Proof of the main result. The existence of at least one continuous reversible interaction implies that we can find $n \geq 2$ qubits such that $\mathcal{G} \neq \mathcal{G}_{\text{loc}}$ and

Theorem 1 holds. Furthermore, as the systems are identical we can perform either G^U or H^U between any pair of qubits.

We now show that the ability to implement one (and thus all) H^U between any two qubits is inconsistent. Note that by acting on three qubits with a sequence of local unitaries, H^U_{12} and H^U_{13} transformations we could implement $(T_1 \circ \mathrm{ad}_V \circ T_1)$ for any $V \in SU(8)$. However, this includes the transformations G^U_{23}. If this were possible, we could first prepare a state ρ_{23} with a negative eigenvalue using H^U_{23} and $\mathcal{G}_{\mathrm{loc}}$ (e.g. by implementing $(T_2 \circ \mathrm{ad}_V \circ T_2)[|00\rangle\langle 00|]$ when $V|00\rangle = \frac{1}{\sqrt{2}}[|00\rangle + |11\rangle]$), then map the negative eigenvector onto the $|00\rangle$ state using G^U_{23} and $\mathcal{G}_{\mathrm{loc}}$. The final state would assign a negative value to the probability of obtaining 00 in a local computational basis measurement, which is inconsistent.

The only remaining possibility is that we can implement G^U between any two qubits. In conjunction with local transformations, this allows us to implement any unitary transformation. Given that we can also perform local preparations and measurements, this allows us to create any quantum state, and to implement any CPTP map or quantum measurement. We now rule out any other states, measurements, and transformations. Note that any state ρ with a negative eigenvalue would assign a negative probabilty for the outcome of some quantum measurement, and is therefore inconsistent. Non-quantum measurement effects can be ruled out similarly. Finally, note that transformations must be completely positive maps, or they could be used to generate a state with a negative eigenvalue by acting on part of some entangled state. As all transformations must be trace-preserving, they must be CPTP maps. Hence the set of allowed states, measurements and transformations is precisely that of standard quantum theory. □

One interesting corollary of our main resul for standard quantum theory is that any continuous reversible interaction, plus local operations and permutations, is universal for quantum computation.The above proof is similar to an argument from [DB11].

To help us prove Theorem 3.1 on which we base the proof of our main result, we first discuss an alternative representation of states based on Bloch-vectors, and the Lie algebra of $\mathcal{G}$. Those who are not interested in the technical details of the proof may move to the Discussion Section.

A multi-qubit Bloch-vector representation

As we are considering qubits, it is helpful to adopt a generalised Bloch vector representation of the state. We expand ρ as

$$\rho = 2^{-n} \sum_{\alpha_1,\ldots,\alpha_n} r_{\alpha_1\cdots\alpha_n}\, \sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n}, \tag{3.0}$$

where $\alpha_k \in \{0,1,2,3\}$ and σ_0 is the identity operator. For clarity, we will use the convention throughout that $\alpha, \beta, \gamma \in \{0,1,2,3\}$ and $i, j \in \{1,2,3\}$.

The real vector

$$r_{\alpha_1\cdots\alpha_n} = \mathrm{tr}\Big((\sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n})\, \rho \Big). \tag{3.0}$$

is a complete representation of the state, and is related to the probability distribution $P(a_1,\ldots,a_n|x_1,\ldots,x_n)$ by an invertible linear map.

For a single qubit $r_0 = 1$ and r_i is the Bloch vector. Similarly, n-qubit product states can be represented by r-vectors of the form

$$r = v(\mathbf{a}_1,\ldots,\mathbf{a}_n) = \begin{bmatrix} 1 \\ \mathbf{a}_1 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ \mathbf{a}_n \end{bmatrix}. \tag{3.0}$$

where the $\mathbf{a}_k$ are Bloch vectors ($\mathbf{a}_k \in \mathbb{R}^3$ and $|\mathbf{a}_k| \leq 1$). As we are considering a locally quantum theory, all these states are elements of the global state space.

Similarly, each measurement outcome can be associated with an *effect* vector p, such that the probability of getting that outcome when measuring the state r is $p^{\mathrm{T}} r$. The vectors $p = 2^{-n} v(\mathbf{b}_1,\ldots,\mathbf{b}_n)$, where the $\mathbf{b}_k$ are Bloch vectors, all correspond to allowed product effects.

Transformations are represented by matrices acting on the state vector. In particular, the transformation $\rho \to G[\rho]$ is represented by the matrix

$$H^{\alpha_1\cdots\alpha_n}_{\beta_1\cdots\beta_n} = \frac{1}{2^n} \mathrm{tr}\Big((\sigma_{\beta_1} \otimes \cdots \otimes \sigma_{\beta_n})\, G[\sigma_{\alpha_1} \otimes \cdots \otimes \sigma_{\alpha_n}] \Big).$$

which acts on the r vector as $r \to Hr$. The single qubit unitaries form a group with a simple matrix representation

$$\mathcal{H}_{\mathrm{q}} = \left\{ \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array}\right] : R \in \mathrm{SO}(3) \right\}. \tag{3.0}$$

We denote the analogues of $\mathcal{G}$ and $\mathcal{G}_{\mathrm{loc}}$ in this representation by $\mathcal{H}$ and $\mathcal{H}_{\mathrm{loc}}$ respectively.

The positivity of probabilities implies that the state space (the set of allowed r's) is bounded. Since $\mathcal{H}$ maps the (bounded) state space into itself, it must be a bounded set of matrices. Without changing the physical predictions of the theory, we can assume that $\mathcal{H}$ is topologically closed [MM11]. It is shown in [Bak02] that all compact matrix groups are Lie groups. Since $\mathcal{H}$ is connected, there is a Lie algebra $\mathfrak{h}$ such that, for each $H \in \mathcal{H}$ there is a matrix $X \in \mathfrak{h}$ satisfying $H = \mathrm{e}^X$ (recall Section 2.1.3 of the Preliminaries Chapter). The quantum Lie algebra is the real vector space of traceless anti-Hermitian matrices, with basis

$$\{\mathrm{i}(\sigma_{\gamma_1} \otimes \ldots \otimes \sigma_{\gamma_n}) \mid (\gamma_1, \ldots, \gamma_n) \neq (0, \ldots, 0)\}.$$

which act on the state through the commutator $\rho \to [\mathrm{i}(\sigma_{\gamma_1} \otimes \cdots \otimes \sigma_{\gamma_n}), \rho]$. When $n = 2$, the matrix representation of any basis element $X \in \mathfrak{h}$ is

$$X^{\alpha_1\alpha_2}_{\beta_1\beta_2} = \frac{1}{2^2}\mathrm{tr}\Big((\sigma_{\beta_1} \otimes \sigma_{\beta_2})\Big[\mathrm{i}(\sigma_{\gamma_1} \otimes \sigma_{\gamma_2}), (\sigma_{\alpha_1} \otimes \sigma_{\alpha_2})\Big]\Big).$$

Defining I as the 4×4 identity matrix, and

$$A_{\mathbf{a}} = \left[\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & a_3 & -a_2 \\ 0 & -a_3 & 0 & a_1 \\ 0 & a_2 & -a_1 & 0 \end{array}\right], \quad B_{\mathbf{a}} = \left[\begin{array}{c|ccc} 0 & a_1 & a_2 & a_3 \\ \hline a_1 & 0 & 0 & 0 \\ a_2 & 0 & 0 & 0 \\ a_3 & 0 & 0 & 0 \end{array}\right], \tag{3.0}$$

we can write the matrix representation of the Lie algebra element $\mathrm{i}(\sigma_i \otimes \sigma_j)$ as

$$X = 2A_{\mathbf{e}_i} \otimes B_{\mathbf{e}_j} + 2B_{\mathbf{e}_i} \otimes A_{\mathbf{e}_j}, \tag{3.0}$$

where $\mathbf{e}_i$ is the unit vector in the i-direction. The element $\mathrm{i}(\sigma_i \otimes \sigma_0)$ has the matrix representation $X = 2A_{\mathbf{e}_i} \otimes I$.

Proof of Theorem 3.1. In the Bloch-representation described above, condition 2 from Theorem 1 is

$$2^{-n} v(\mathbf{b}_1, \ldots, \mathbf{b}_n)^{\mathrm{T}} H v(\mathbf{a}_1, \ldots, \mathbf{a}_n) \in [0, 1], \tag{3.0}$$

for all Bloch vectors $\mathbf{a}_r, \mathbf{b}_r$. Considering a group element close to the identity, $H = \mathrm{e}^{\epsilon X} \in \mathcal{H}$, and expanding equation (B.-17) to second order in ϵ gives

$$v(\mathbf{b}_1 \ldots \mathbf{b}_n)^{\mathrm{T}}\left(I^{\otimes n} + \epsilon X + \frac{\epsilon^2}{2}X^2\right) v(\mathbf{a}_1 \ldots \mathbf{a}_n) \in [0, 2^n] \tag{3.0}$$

When all the Bloch vectors have unit length, we can use this expansion to derive the first-order constraints

$$\mathcal{C}[\mathbf{a}_1] \equiv v(-\mathbf{a}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)^{\mathrm{T}} X v(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = 0, \tag{3.0}$$

and the second-order constraints

$$\begin{aligned} v(-\mathbf{a}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n)^{\mathrm{T}} X^2 v(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) &\geq 0, & (3.1)\\ v(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)^{\mathrm{T}} X^2 v(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) &\leq 0. & (3.2) \end{aligned}$$

These constraints hold for all $X \in \mathfrak{h}$.

We initially use the first-order constraints in (3.2.2). Considering $\mathcal{C}[\mathbf{e}_i] \pm \mathcal{C}[-\mathbf{e}_i] = 0$, we find

$$X^{i\alpha_2\cdots\alpha_n}_{i\beta_2\cdots\beta_n} = X^{0\alpha_2\cdots\alpha_n}_{0\beta_2\cdots\beta_n} \quad \text{and} \quad X^{i\alpha_2\cdots\alpha_n}_{0\beta_2\cdots\beta_n} = X^{0\alpha_2\cdots\alpha_n}_{i\beta_2\cdots\beta_n}. \tag{3.2}$$

for all values of i, α_r, β_r, where we have used the fact that the vectors $v(\mathbf{a}_2, \cdots, \mathbf{a}_n)$ linearly span the whole of $(\mathbb{R}^4)^{\otimes(n-1)}$.

Similarly, taking $\mathcal{C}[\frac{1}{\sqrt{2}}(\mathbf{e}_i + \mathbf{e}_j)] + \mathcal{C}[-\frac{1}{\sqrt{2}}(\mathbf{e}_i + \mathbf{e}_j)] = 0$ for $i \neq j$, and using (3.2.2), we obtain

$$X^{i\alpha_2\cdots\alpha_n}_{j\beta_2\cdots\beta_n} = -X^{j\alpha_2\cdots\alpha_n}_{i\beta_2\cdots\beta_n}. \tag{3.2}$$

Let $\mathcal{A}$ and $\mathcal{B}$ denote the linear spans of the $A_{\mathbf{a}}$ and $B_{\mathbf{a}}$ matrices defined in (3.2.2) respectively, and let $\mathcal{I}$ denote the linear span of I. Then equations (3.2.2-3.2.2), together with the equivalent equations for the other $(n-1)$ qubits, imply that $\mathfrak{h}$ is a subspace of $(\mathcal{A} \oplus \mathcal{B} \oplus \mathcal{I})^{\otimes n}$. We equip $(\mathcal{A} \oplus \mathcal{B} \oplus \mathcal{I})$ with the standard matrix inner-product $\langle A, B\rangle = \mathrm{tr}(A^{\mathrm{T}}B)$, under which $\{A_{\mathbf{e}_1}, A_{\mathbf{e}_2}, A_{\mathbf{e}_3}, B_{\mathbf{e}_1}, B_{\mathbf{e}_2}, B_{\mathbf{e}_3}, I\}$ form an orthogonal basis for the space.

It is easy to see that the local Lie algebra is

$$\mathfrak{h}_{\mathrm{loc}} = \bigoplus_{\pi} \pi(\mathcal{A} \otimes \mathcal{I}^{\otimes(n-1)}), \tag{3.2}$$

where π runs over all permutations of the n factor spaces. Condition 1 from Theorem 1 implies $\mathfrak{h}_{\mathrm{loc}} \subseteq \mathfrak{h}$.

If $X \notin \mathfrak{h}_{\mathrm{loc}}$ then we can always re-order the subsystems such that X has non-zero overlap with the subspace $\mathcal{S} = \mathcal{A}^{\otimes n_A} \otimes \mathcal{B}^{\otimes n_B} \otimes \mathcal{I}^{\otimes n_I}$, where $n_A + n_B + n_I = n$, and at least one of the inequalities, $n_A \geq 2$ or $n_B \geq 1$, holds. In particular, there is a matrix $M_S = A_{\mathbf{a}_1} \otimes \ldots \otimes A_{\mathbf{a}_{n_A}} \otimes B_{\mathbf{b}_1} \otimes \ldots \otimes B_{\mathbf{b}_{n_B}} \otimes I^{\otimes n_I} \in \mathcal{S}$ that has non-zero overlap with X. Furthermore, there exists a local transformation $H_{\mathrm{loc}} \in \mathcal{H}$ such that $H_{\mathrm{loc}} M_S H_{\mathrm{loc}}^{-1} \propto A_{\mathbf{e}_1}^{\otimes n_A} \otimes B_{\mathbf{e}_1}^{\otimes n_B} \otimes I^{\otimes n_I}$.

Let us define $E_0 = A_{\mathbf{e}_1}$ and $E_1 = B_{\mathbf{e}_1}$, and denote the linear span of these two matrices by $\mathcal{E}$. Then $H_{\rm loc} X H_{\rm loc}^{-1}$ has support on $\mathcal{E}^{\otimes m} \otimes \mathcal{I}^{\otimes n_I}$, where $m = n_A + n_B$.

For any matrix $M \in (\mathcal{A} \oplus \mathcal{B} \oplus \mathcal{I})$, the projector onto $\mathcal{I}$ is given by $\Phi_{\mathcal{I}}[M] = \int_{\mathcal{H}_q} dH\, HMH^{-1}$, where $\mathcal{H}_q$ is the group of all single qubit unitaries defined in (3.2.2). Similarly, the projector onto $\mathcal{E}$ is given by $\Phi_{\mathcal{E}}[M] = \int_{\mathcal{H}_{\mathbf{e}_1}} dH\, HMH^{-1} - \Phi_{\mathcal{I}}[M]$ where $\mathcal{H}_{\mathbf{e}_1} = \{H \in \mathcal{H}_q; Hv(\mathbf{e}_1) = v(\mathbf{e}_1)\}$. As $HXH^{-1} \in \mathfrak{h}$ whenever $X \in \mathfrak{h}$ and $H \in \mathcal{H}$, it follows that the matrix

$$Y = (\Phi_{\mathcal{E}}^{\otimes m} \otimes \Phi_{\mathcal{I}}^{\otimes n_I})\big[H_{\rm loc} X H_{\rm loc}^{-1}\big] \tag{3.2}$$

is a non-zero element of $\mathfrak{h}$, and fully contained in $\mathcal{E}^{\otimes m} \otimes \mathcal{I}^{\otimes n_I}$. This allows us to expand Y as

$$Y = \sum_{\mathbf{s}\in\{0,1\}^m} c_{\mathbf{s}}\; E_{s_1} \otimes \cdots \otimes E_{s_m} \otimes I^{\otimes n_I}, \tag{3.2}$$

Since $E_0 E_1 = E_1 E_0 = 0$ we have

$$Y^2 = \sum_{\mathbf{s}\in\{0,1\}^m} c_{\mathbf{s}}^2\; E_{s_1}^2 \otimes \cdots \otimes E_{s_m}^2 \otimes I^{\otimes n_I}. \tag{3.2}$$

Also note that

$$v(\pm\mathbf{e}_2)^{\rm T} E_0^2\, v(\mathbf{e}_2) = \mp 1, \qquad v(\pm\mathbf{e}_2)^{\rm T} E_1^2\, v(\mathbf{e}_2) = 1, \tag{3.3}$$

$$v(\mathbf{e}_1)^{\rm T} E_0^2\, v(\mathbf{e}_1) = 0, \qquad v(\mathbf{e}_1)^{\rm T} E_1^2\, v(\mathbf{e}_1) = 2. \tag{3.4}$$

We now use these equations, together with the second order constraints given by (3.1) and (3.2), to derive some properties of the coefficients $c_{\mathbf{s}}$. First note that according to (3.2)

$$v(\mathbf{e}_1, \ldots, \mathbf{e}_1)^{\rm T} Y^2 v(\mathbf{e}_1, \ldots, \mathbf{e}_1) = c_{(1,1,1,\ldots,1)}^2\, 2^n \leq 0 \tag{3.4}$$

hence $c_{(1,1,1,\ldots,1)} = 0$. This also rules out the case $n_A = 0, n_B = 1$, which implies $m \geq 2$.

Now consider the two inequalities

$$v(-\mathbf{e}_2, \mathbf{e}_2, \mathbf{e}_1, \ldots, \mathbf{e}_1)^{\rm T} Y^2 v(\mathbf{e}_2, \mathbf{e}_2, \mathbf{e}_1, \ldots, \mathbf{e}_1) \geq 0, \tag{3.5}$$

$$v(\mathbf{e}_2, -\mathbf{e}_2, \mathbf{e}_1, \ldots, \mathbf{e}_1)^{\rm T} Y^2 v(\mathbf{e}_2, \mathbf{e}_2, \mathbf{e}_1, \ldots, \mathbf{e}_1) \geq 0, \tag{3.6}$$

derived from (3.1). Adding these two inequalities together we obtain $(c_{(1,1,1,\ldots,1)}^2 - c_{(0,0,1,\ldots,1)}^2) 2^{n-1} \geq 0$, which implies $c_{(0,0,1,\ldots,1)} = 0$. After removing the terms which are zero in these two inequalities we obtain

$$(c_{(0,1,1,\ldots,1)}^2 - c_{(1,0,1,\ldots,1)}^2) 2^{n-2} \geq 0,$$
$$(c_{(1,0,1,\ldots,1)}^2 - c_{(0,1,1,\ldots,1)}^2) 2^{n-2} \geq 0,$$

respectively. Together these imply $c_{(1,0,1,\ldots,1)} = \pm c_{(0,1,\ldots,1)}$.

We now show by induction that $c_{(0,1,\ldots,1)} \neq 0$. From the construction of Y, it is clear that $c_{(0\ldots0,1,\ldots1)} \neq 0$ when the index contains n_A zeroes. Now, for some $l \geq 2$, suppose that $c_{(s_1,\ldots,s_l,1,\ldots,1)} \neq 0$ if and only if $s_1 = \cdots = s_l = 0$. In this case, the constraint

$$v(\pm\mathbf{e}_2, -\mathbf{e}_2 \ldots - \mathbf{e}_2, \mathbf{e}_1 \ldots \mathbf{e}_1)^{\mathrm{T}} Y^2 v(\mathbf{e}_2 \ldots \mathbf{e}_2, \mathbf{e}_1 \ldots \mathbf{e}_1) \geq 0$$

implies $\mp c^2_{(0\ldots0,1\ldots1)} 2^{n-l} \geq 0$, which is impossible. Hence, there must exist another component $c_{(s_1,\ldots,s_l,1,\ldots,1)} \neq 0$ for which the index contains less than l zeroes. Rearranging the first l qubits such that the zeroes occur at the start of $\mathbf{s}$ and proceeding by induction, we find that either $c_{(0,1,\ldots,1)} \neq 0$ or $c_{(1,1,\ldots,1)} \neq 0$. However, the latter possibility is ruled out above.

Taking the $n-2$ leftmost qubits to be in the state $\sigma = v(\mathbf{e}_1)$, and noting that $E_0\sigma = 0$ and $E_1\sigma = \sigma$, the element $Z = c^{-1}_{(0,1,1,\ldots,1)} Y \in \mathfrak{h}$ acts as

$$Z\Big(\mathbf{r}_{12} \otimes \sigma^{\otimes(n-2)}\Big) = \big((A_{\mathbf{e}_1} \otimes B_{\mathbf{e}_1} \pm B_{\mathbf{e}_1} \otimes A_{\mathbf{e}_1})\mathbf{r}_{12}\big) \otimes \sigma^{\otimes(n-2)}$$

for any vector $\mathbf{r}_{12}$. In the $+$ case, the action on the first two qubits is that of the quantum Lie algebra element $i(\sigma_1 \otimes \sigma_1)$, which can be used to generate the entangling unitary transformation $U = e^{i(\sigma_1 \otimes \sigma_1)}$. In the $-$ case, the Lie algebra element on the first two qubits can be written as $-T_1(A_{\mathbf{e}_1} \otimes B_{\mathbf{e}_1} + B_{\mathbf{e}_1} \otimes A_{\mathbf{e}_1})T_1$, where T_1 is the transpose operation on the first qubit. This can be used to generate the transformation $T_1 \circ \mathrm{ad}_U \circ T_1$. Rewriting these results in terms of $G[\rho]$, we recover Theorem 1. □

3.3. Discussion

In this Chapter we have shown that quantum theory is the only theory in which (i) local systems behave like identical qubits (ii) there exists at least one continuous, reversible interaction. This highlights the importance of dynamical reversibility to the non-local structure of quantum theory.

Our results also have implications for the viability of large scale quantum computation. It is conceivable that quantum theory could experience a fundamental breakdown once a particular threshold in the number of qubits (say, millions) has been achieved. Our results make this possibility more unlikely since, if so, one of our assumptions must be violated: either local quantumness, or continuous reversibility, or local tomography must fail. These violations could then be addressed experimentally, without necessarily having to test this on millions of qubits.

Because we use Bloch-spheres and Lie algebras in our proof, it does not easily generalise to higher dimensional systems or discrete transformations. However, we conjecture that quantum theory is the only theory which is locally quantum, and in which there exists a reversible interaction between any pair of systems. More generally, an interesting open question is whether all reversible, locally-tomographic theories can be represented within quantum theory.

4. Full randomness amplification

Do completely unpredictable events exist? Classical physics excludes intrinsic randomness. While quantum theory makes probabilistic predictions, this does not imply that nature is random, as randomness should be certified without relying on the complete structure of the theory being used. As we saw in the Preliminaries Chapter, Section 2.2, Bell tests approach the question from this device-independent perspective. However, they require prior perfect randomness, falling into a circular reasoning. A Bell test that generates perfect random bits from bits possessing high -but less than perfect- randomness was obtained in [CR12b]. Yet, the main question remained open: do sources of arbitrarily weak randomness suffice to certify perfect randomness?

In this Chapter we provide an affirmative answer to that question providing the first protocol for *full randomness amplification* a task known to be impossible using classical resources. The results of this Chapter are based on the article [GMDLT+13].

4.1. Introduction

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

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

The advent of quantum physics put into question this deterministic viewpoint, as there exist experimental situations for which quantum theory gives predictions only in probabilistic terms, even if one has a perfect description of

the preparation and interactions of the system. A possible solution to this classically counterintuitive fact was proposed in the early days of quantum physics: Quantum mechanics had to be incomplete [EPR35], and there should be a complete theory capable of providing deterministic predictions for all conceivable experiments. There would thus be no room for intrinsic randomness, and any apparent randomness would again be a consequence of our lack of control over hypothetical "hidden variables" not contemplated by the quantum formalism.

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

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

The main result of this Chapter is that this is the case for a very general, and physically-meaningful, set of randomness sources. This includes subsets of the well known Santha-Vazirani sources [SV86] as particular cases. Besides

the philosophical and physics-foundational implications, our results provide a protocol for full randomness amplification using quantum non-locality. Randomness amplification is an information-theoretic task whose goal is to use an input source of imperfectly random bits to produce perfect random bits. Santha and Vazirani proved that randomness amplification is impossible using classical resources [SV86]. This is in a sense intuitive, in view of the absence of any intrinsic randomness in classical physics. In the quantum regime, randomness amplification has been recently studied by Colbeck and Renner [CR12b]. They proved how input bits with very high initial randomness can be mapped into arbitrarily pure random bits, and conjectured that randomness amplification should be possible for any initial randomness [CR12b]. Our results also solve this conjecture, as we show that quantum non-locality can be exploited to attain full randomness amplification.

4.2. Results

4.2.1. Previous work

Before presenting our results, it is worth commenting on previous works on randomness in connection with quantum non-locality. In [PAM+10b] it was shown how to bound the intrinsic randomness generated in a Bell test. These bounds can be used for device-independent randomness expansion, following a proposal by Colbeck [Col07], and to achieve a quadratic expansion of the amount of random bits (see [AMP12a, PM11, FGS11, VV12a] for further works on device-independent randomness expansion). Note however that, in randomness expansion, one assumes instead, from the very beginning, the existence of an input seed of free random bits, and the main goal is to expand this into a larger sequence. The figure of merit there is the ratio between the length of the final and initial strings of free random bits. Finally, other recent works have analyzed how a lack of randomness in the measurement choices affects a Bell test [KPB06, BG10, Hal10] and the randomness generated in it [KHS+12].

4.2.2. Definition of the scenario

From an information perspective, our goal is to construct a protocol for full randomness amplification based on quantum non-locality. In randomness amplification, one aims at producing arbitrarily free random bits from many uses of an input source $\mathcal{S}$ of imperfectly random bits.

A random bit b is said to be free if it is uncorrelated from any classical variables e generated outside the future light-cone of b (of course, the bit b can

be arbitrarily correlated with any event inside its future light-cone). This requirement formalises the intuition that the only systems that may share some correlation with b are the ones that are influenced by b. Note also that this definition of randomness is strictly stronger than the demand that b is uncorrelated with any classical variable generated in the past light-cone of the process. This is crucial if the variable e and b are generated by measuring on a correlated quantum system. In this case, even if both systems interacted somewhere in the past light-cone of b, the variable e is not produced until the measurement is performed, possibly outside both past and future light-cones. Furthermore, we say that a random bit is ϵ-free if any correlations with events outside its future light-cone are bounded by ϵ, as explained in what follows.

Source $\mathcal{S}$ produces a sequence of bits $x_1, x_2, \ldots x_j, \ldots$, with $x_j = 0$ or 1 for all j, see Fig. 4.1, which are ϵ-free. More precisely, each bit j contains some randomness, in the sense that the probability $P(x_j|\text{all other bits}, e)$ that it takes a given value x_j, conditioned on the values of all the other bits produced by $\mathcal{S}$, as well as the variable e, is such that

$$\epsilon \leq P(x_j|\text{all other bits}, e) \leq 1 - \epsilon \tag{4.0}$$

for all j, where ϵ takes a fixed value in the range $0 < \epsilon \leq 1/2$. Given our previous definition of ϵ-free bits, the variable e represents events outside the future light-cone of all the x_j's. Free random bits correspond to $\epsilon = \frac{1}{2}$; while deterministic ones to $\epsilon = 0$. More precisely, when $\epsilon = 0$ the bound (4.2.2) is trivial and no randomness can be certified. We refer to $\mathcal{S}$ as an ϵ-source, and to any bit satisfying (4.2.2) as an ϵ-free bit.

The aim of randomness amplification is to generate, from arbitrarily many uses of $\mathcal{S}$, a final source $\mathcal{S}_f$ of ϵ_f arbitrarily close to 1/2. If this is possible, no cause e can be held responsible (no even partially) of the bits produced by $\mathcal{S}_f$, which are then fully unpredictable. Note that, in our case, we require the final bits to be fully uncorrelated from e. When studying randomness amplification, the first and most fundamental question is whether the process is at all possible. This is the question we consider and solve in this Chapter. Thus, we are not concerned with efficiency issues, such as the rate of uses of $\mathcal{S}$ required per final bit generated by $\mathcal{S}_f$, and, without loss of generality, restrict our analysis to the problem of generating a single final free random bit k. The relevant figure of merit in this Chapter is just the quality, measured by ϵ_f, of the final bit. Of course, efficiency issues are relevant when considering applications of randomness amplification protocols for information tasks, but this is beyond the scope of the Chapter.

The randomness amplification protocols we consider exploit quantum nonlocality. This idea was introduced in [CR12b], where a protocol was presented

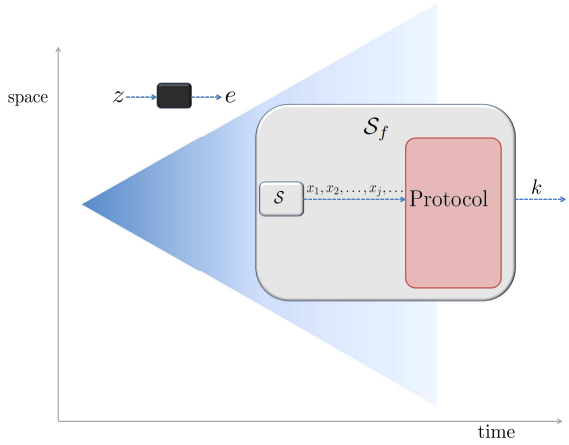

Figure 4.1.: Local causal structure and randomness amplification. A source $\mathcal{S}$ produces a sequence $x_1, x_2, \ldots x_j, \ldots$ of imperfect random bits. The goal of randomness amplification is to produce a new source $\mathcal{S}_f$ of perfect random bits, that is, to process the initial bits so as to get a final bit k fully uncorrelated (free) from all events outside the future light cone of all the bits x_j produced by the source. In other words k is free if it is uncorrelated from any event outside the lightcone shown in this figure. Any such event can be modeled by a measurement z, with an outcome e, on some physical system. This system may be under the control of an adversary Eve, interested in predicting the value of k.

in which the source $\mathcal{S}$ is used to choose the measurement settings by two distant observers, Alice and Bob, in a Bell test [BC90] involving two entangled quantum particles. The measurement outcome obtained by one of the observers, say Alice, in one of the experimental runs (also chosen with $\mathcal{S}$) defines the output random bit. Colbeck and Renner proved how input bits with high randomness, of $0.442 < \epsilon \leq 0.5$, can be mapped into arbitrarily free random bits of $\epsilon_f \to 1/2$. In our case, the input ϵ-source $\mathcal{S}$ is used to choose the measurement settings in a multipartite Bell test involving a number of observers that depends both on the input ϵ and the target ϵ_f. After verifying that the expected Bell violation is obtained, the measurement outcomes are combined to define the final bit k. For pedagogical reasons, we adopt a cryptographic perspective and assume the worst-case scenario where all the devices we use may have been prepared by an adversary Eve equipped with arbitrary non-signaling resources, possibly even supra-quantum ones. In the preparation, Eve may have also had access to $\mathcal{S}$

and correlated the bits it produces with some physical system at her disposal, represented by a black box in Fig. 4.1. Without loss of generality, we can assume that Eve can reveal the value of e at any stage of the protocol by measuring this system. Full randomness amplification is then equivalent to proving that Eve's correlations with k can be made arbitrarily small.

An important comment is now in order that applies to all further discussion as well as the protocol subsequently presented. For convenience we represent (see Figs. 4.1 and 4.2) $\mathcal{S}$ as a single source generating all the inputs and delivering them among the separated boxes without violating the no-signalling principle. However, this is not the scenario in practice. Operationally, each user generates his input from a local source in his lab. However, the sources of all users can be arbitrarily correlated with each other, without violating the bound on the correlations given by (4.2.2), and thus, can be seen as a single ϵ-source $\mathcal{S}$. With this understanding we proceed to discuss a single effective source in the rest of the text.

4.2.3. Partial randomness from GHZ-type paradoxes

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

Let us give an explicit attack for the CHSH inequality [CHSH69] to illustrate this point. As we saw in the preliminaries Chapter, the inequality can be written as $B(P_{\text{obs}}) \equiv \langle A_0B_0\rangle + \langle A_1B_0\rangle + \langle A_0B_1\rangle - \langle A_1B_1\rangle$. The algebraic value of the inequality is 4 when the first three correlators take value $+1$ and the last correlator takes value -1. There are classical models which assign the mentioned values to all correlators but one, to which they assign the opposite value. This accounts to a total CHSH value of 2 which is the classical bound. Let us consider four classical models $\lambda_{A_iB_j}$, labeled by the correlator in which they give the value opposite to the one needed for the algebraic violation. Imagine Eve draws each model with probability $p(\lambda_{A_iB_j}) = 1/4 \;\forall\; i,j$. Conditioned on a particular model, Eve can fix what inputs x, y will be drawn by Alice and Bob up to epsilon for each of the bits. Hence, the expected value of each correlator is $\langle A_iB_j\rangle = +1\cdot(1-\epsilon^2) - 1\cdot\epsilon^2 = 1-2\epsilon^2$ for the first three correlators, and the same up to an overall negative sign for the last correlator. The expected violation with this attack is then $B(P_{\text{obs}}) = 4(1-2\epsilon^2)$. One can see that, for all but the algebraic violation violation $B(P_{\text{obs}}) = 4$, there exist classical models

were parties using non-zero randomness $1/2 \geq \epsilon > 0$ can nonetheless observe a fake violation. In fact, it is enough to loose about 23% of 'free will' as captured by ϵ so as to obtain the maximal quantum violation for this inequality $B(P_{\text{obs}}) = 2\sqrt{2}$. We hence need inequalities that admit a maximal violation within quantum mechanics, those are called GHZ-type paradoxes.

Nevertheless, GHZ-type paradoxes are not sufficient. Consider for instance the standard 3−qubit GHZ paradox. We later show that given any function of the measurement outcomes, it is always possible to find non-signaling correlations that (i) maximally violate the 3-party GHZ paradox[GHZ89] but (ii) assigns a deterministic value to that function of the measurement outcomes. This observation can be checked for all unbiased functions mapping $\{0,1\}^3$ to $\{0,1\}$ (there are $\binom{8}{4}$ of those) through a linear program analogous to the one used in the proof of the Lemma 4.1 stated below. As a simple example, consider the particular function defined by the outcome bit of the first user. This can be fixed by using a tripartite no-signaling probability distribution consisting on a deterministic distribution for the first party, and a PR box [PR94] for the second and third party. As a second and more complicated example, imagine that the function is instead the parity of all three outcomes $a_1 \oplus a_2 \oplus a_3$. It is easy to see that via the same tripartite no-signaling probability distribution the outcome of such function can be known to Eve. This is true since the PR box has its parity fixed once the inputs are known and the other party is deterministic as we said.

For five parties though, the latter happens not to hold any longer. Consider now any correlations attaining the maximal violation of the five-party Mermin inequality [Mer90]. In each run of this Bell test, measurements (inputs) $\mathbf{x} = (x_1, \ldots, x_5)$ on five distant black boxes generate 5 outcomes (outputs) $\mathbf{a} = (a_1, \ldots, a_5)$, distributed according to a non-signaling conditional probability distribution $P(\mathbf{a}|\mathbf{x})$, see Appendix B. Both inputs and outputs are bits, as they can take two possible values, $x_i, a_i \in \{0,1\}$ with $i = 1, \ldots, 5$.

The inequality can be written as

$$\sum_{\mathbf{a},\mathbf{x}} I(\mathbf{a},\mathbf{x}) P(\mathbf{a}|\mathbf{x}) \geq 6 \ , \tag{4.0}$$

with coefficients

$$I(\mathbf{a},\mathbf{x}) = (a_1 \oplus a_2 \oplus a_3 \oplus a_4 \oplus a_5)\, \delta_{\mathbf{x}\in\mathcal{X}_0} + (a_1 \oplus a_2 \oplus a_3 \oplus a_4 \oplus a_5 \oplus 1)\, \delta_{\mathbf{x}\in\mathcal{X}_1} \ , \tag{4.0}$$

where

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

$$\mathcal{X}_0 = \Big\{\mathbf{x} \;\Big|\; \sum_{i=1}^{5} x_i = 1\Big\} \cup \Big\{\mathbf{x} \;\Big|\; \sum_{i=1}^{5} x_i = 5\Big\}, \tag{4.0}$$

and

$$\mathcal{X}_1 = \Big\{\mathbf{x} \;\Big|\; \sum_{i=1}^{5} x_i = 3\Big\}. \tag{4.0}$$

That is, only half of all possible combinations of inputs, namely those in $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$, appear in the Bell inequality. This inequality may be thought of as a non-local game in which the parties are required to minimize the parity of their outputs when the sum of their inputs is 1 or 5 while minimizing the inverse parity of the outputs when their inputs sum to 3. It turns out that the minimum achievable with classical strategies is 6.

The maximal, non-signaling and algebraic, violation of the inequality corresponds to the situation in which the left-hand side of (4.2.3) is zero. The key property of inequality (4.2.3) is that its maximal violation can be attained by quantum correlations and furthermore, one can construct a function of the outcomes that is not completely determined. Take the bit corresponding to the majority-vote function of the outcomes of any subset of three out of the five observers, say the first three. This function is equal to zero if at least two of the three bits are equal to zero, and equal to one otherwise. We show that Eve's predictability on this bit is at most 3/4. We state this result in the following Lemma:

Lemma 4.1. *Let a five-party non-signaling conditional probability distribution $P(\mathbf{a}|\boldsymbol{x})$ in which inputs $\boldsymbol{x} = (x_1, \ldots, x_5)$ and outputs $\mathbf{a} = (a_1, \ldots, a_5)$ are bits. Consider the bit* $\mathrm{maj}(\mathbf{a}) \in \{0, 1\}$ *defined by the majority-vote function of any subset consisting of three of the five measurement outcomes, say the first three, a_1, a_2 and a_3. Then, all non-signaling correlations attaining the maximal violation of the 5-party Mermin inequality are such that the probability that* $\mathrm{maj}(\mathbf{a})$ *takes a given value, say 0, is bounded by*

$$1/4 \leq P\left(\mathrm{maj}(\mathbf{a}) = 0\right) \leq 3/4. \tag{4.0}$$

Proof of lemma 4.1. This result was obtained by solving a linear program. Therefore, the proof is numeric, but exact. Formally, let $P(\mathbf{a}|\mathbf{x})$ be a 5-partite non-signaling probability distribution. For $\mathbf{x} = \mathbf{x}_0 \in \mathcal{X}$, we performed the

maximization,

$$\begin{aligned} P_{max} = \max_{P} \; & P(\mathrm{maj}(\mathbf{a}) = 0|\mathbf{x}_0) \\ & \text{subject to} \\ & I(\mathbf{a}, \mathbf{x}) \cdot P(\mathbf{a}|\mathbf{x}) = 0 \end{aligned} \tag{4.0}$$

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

As a further remark, note that a lower bound to P_{max} can easily be obtained by noticing that one can construct conditional probability distributions $P(\mathbf{a}|\mathbf{x})$ that maximally violate 5-partite Mermin inequality (4.2.3) for which at most one of the output bits (say a_1) is deterministically fixed to either 0 or 1. If the other two output bits (a_2, a_3) were to be completely random, the majority-vote of the three of them $\mathrm{maj}(a_1, a_2, a_3)$ could be guessed with a probability of 3/4. Our numerical results say that this turns out to be an optimal strategy.

The previous lemma strongly suggests that, given an ϵ-source with any $0 < \epsilon \leq 1/2$ and quantum five-party non-local resources, it should be possible to design a protocol to obtain an ϵ_i-source of $\epsilon_i = 1/4$. We do not explore this possibility here, but rather use the partial unpredictability in the five-party Mermin Bell test as building block of our protocol for full randomness amplification. To complete it, we must equip it with two essential components: (*i*) an estimation procedure that verifies that the untrusted devices do yield the required Bell violation; and (*ii*) a distillation procedure that, from sufficiently many ϵ_i-bits generated in the 5-party Bell experiment, distills a single final ϵ_f-source of $\epsilon_f \to 1/2$. Towards these ends, we consider a more complex Bell test involving N groups of five observers (quintuplets) each.

4.2.4. A protocol for full randomness amplification

Our protocol for randomness amplification uses as resources the ϵ-source $\mathcal{S}$ and $5N$ quantum systems. Each of the quantum systems is abstractly modeled by a black box with binary input x and output a. The protocol processes classically the bits generated by $\mathcal{S}$ and by the quantum boxes. When the protocol is not aborted it produces a bit k. The protocol consists of the 5 steps described below (see also Fig. 4.2).

- **Step 1:** $\mathcal{S}$ is used to generate N quintuple-bits $\mathbf{x}_1, \ldots \mathbf{x}_N$, which constitute the inputs for the $5N$ boxes and are distributed among them. The boxes then provide N output quintuple-bits $\mathbf{a}_1, \ldots \mathbf{a}_N$ without violating no-signaling.

- **Step 2:** The quintuplets such that $\mathbf{x} \notin \mathcal{X}$ are discarded. The protocol is aborted if the number of remaining quintuplets is less than $N/3$. (Note that the constant factor $1/3$ is arbitrary. In fact, it is enough to demand that the number of remaining quintuplets is larger than N/c, with $c > 1$. See the Appendix B).

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

- **Step 4:** The function

$$r[b, y] = \begin{cases} 1 & \text{if } I(\mathbf{a}_1, \mathbf{x}_1) = \dots = I(\mathbf{a}_{N_d}, \mathbf{x}_{N_d}) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.0}$$

  tells whether block (b, y) features the right correlations ($r = 1$) or the wrong ones ($r = 0$), in the sense of being compatible with the maximal violation of inequality (4.2.3). This function is computed for all blocks but the distilling one. The protocol is aborted unless all of them give the right correlations,

$$g = \prod_{j=1, j\neq l}^{N_b} r[b_j, y_j] = \begin{cases} 1 & \text{not abort} \\ 0 & \text{abort} \end{cases} . \tag{4.0}$$

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

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

$$k = f(\mathrm{maj}(\mathbf{a}_1), \dots \mathrm{maj}(\mathbf{a}_{N_d})) . \tag{4.0}$$

  Here $f : \{0,1\}^{N_d} \to \{0,1\}$ is a function whose existence is proven in Appendix B, while $\mathrm{maj}(\mathbf{a}_i) \in \{0,1\}$ is the majority-vote among the three first bits of the quintuple string $\mathbf{a}_i$.

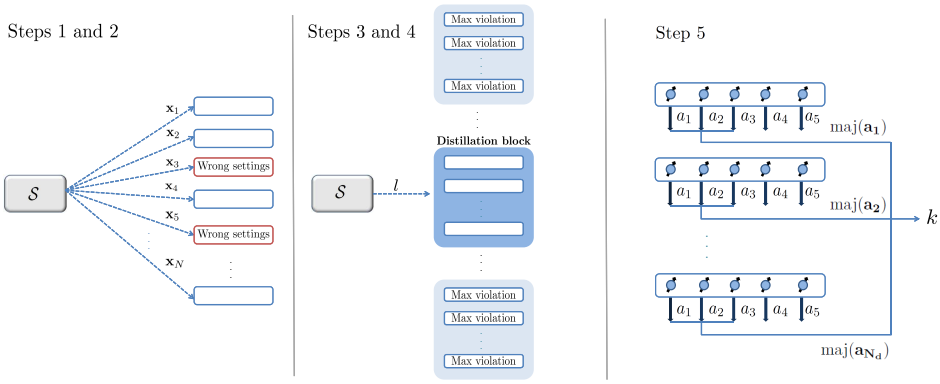

Figure 4.2.: Protocol for full randomness amplification based on quantum non-locality. In the first two steps, all N quintuplets measure their devices, where the choice of measurement is done using the ϵ-source $\mathcal{S}$. Although it is illustrated here as a single source for convenience we recall that it represents the collection of sources that each space-like separated party locally possesses with all their outputs being correlated to form an ϵ-source. The quintuplets whose settings happen not to take place in the five-party Mermin inequality are discarded (in red). In steps 3 and 4, the remaining quintuplets are grouped into blocks. One of the blocks is chosen as the distillation block, using again $\mathcal{S}$, while the others are used to check the Bell violation. In the fifth step, the random bit k is extracted from the distillation block.

At the end of the protocol, the bit k is potentially correlated with the settings of the distilling block $\tilde{y} = y_l$, the bit g defined in (4.2.4), and the information

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

Additionally, an eavesdropper Eve might have access to a physical system correlated with k, which she can measure at any stage of the protocol. This

system is not necessarily classical nor quantum, the only assumption about it is that measuring it does not produce instantaneous signaling anywhere else. The measurements that Eve can perform on her system are labeled by z, and the corresponding outcomes by e. In summary, after performing the protocol all the relevant information is $k, \tilde{y}, t, g, e, z$, with statistics described by an unknown conditional probability distribution $P(k, \tilde{y}, t, g, e|z)$. When the protocol is aborted ($g = 0$) there is no value for k. Therefore, in order to have a well defined distribution $P(k, \tilde{y}, t, g, e|z)$ in all cases, we set $k = 0$ when $g = 0$—that is $P(k, \tilde{y}, t, g = 0, e|z) = \delta^0_k P(\tilde{y}, t, g = 0, e|z)$, where δ^0_k is the Kronecker tensor.

To assess the quality of our protocol for full randomness amplification we compare it with an ideal protocol having the same marginal for the variables $\tilde{y}, t, g$ and the physical system described by e, z. That is, the global distribution of the ideal protocol is

$$P_{\text{ideal}}(k, \tilde{y}, t, g, e|z) = \begin{cases} \frac{1}{2} P(\tilde{y}, t, g, e|z) & \text{if } g = 1 \\ \delta^0_k P(\tilde{y}, t, g, e|z) & \text{if } g = 0 \end{cases} , \tag{4.0}$$

where $P(\tilde{y}, t, g, e|z)$ is the marginal of the distribution $P(k, \tilde{y}, t, g, e|z)$ generated by the real protocol. Note that, consistently, in the ideal distribution we also set $k = 0$ when $g = 0$.

Our goal is that the statistics of the real protocol P are indistinguishable from the ideal statistics P_{ideal}. We consider the optimal strategy to discriminate between P and P_{ideal}, which obviously involves having access to all possible information $k, \tilde{y}, t, g$ and the physical system e, z. As shown in [Mas09], the optimal probability for correctly guessing between these two distributions is

$$P(\text{guess}) = \frac{1}{2} + \frac{1}{4} \sum_{k,\tilde{y},t,g} \max_z \sum_e \left| P(k, \tilde{y}, t, g, e|z) - P_{\text{ideal}}(k, \tilde{y}, t, g, e|z) \right| . \tag{4.0}$$

Note that the second term can be understood as (one half of) the variational distance between P and P_{ideal} generalized to the case when the distributions are conditioned on an input z. The following theorem is proven in Appendix B.

Theorem 4.2 (Full randomness amplification is possible)**.** *Let $P(k, \tilde{y}, t, g, e|z)$ be the probability distribution of the variables generated during the protocol and the adversary's physical system e, z; and let $P_{\text{ideal}}(k, \tilde{y}, t, g, e|z)$ be the corresponding ideal distribution (4.2.4). The optimal probability of correctly guessing between the distributions P and P_{ideal} satisfies*

$$P(\text{guess}) \;\leq\; \frac{1}{2} + \frac{3\sqrt{N_d}}{2} \left[\alpha^{N_d} + 2\, N_b^{\log_2(1-\epsilon)} \left(32 \beta \epsilon^{-5} \right)^{N_d} \right] , \tag{4.0}$$

where the real numbers α, β *fulfill* $0 < \alpha < 1 < \beta$.

Now, the right-hand side of (4.2) can be made arbitrary close to 1/2, for instance by setting $N_b = \left(32\,\beta\,\epsilon^{-5}\right)^{2N_d/|\log_2(1-\epsilon)|}$ and increasing N_d subject to the condition $N_d N_b > N/3$. [Note that $\log_2(1-\epsilon) < 0$.] In the limit of large N_d the probability $P(\text{guess})$ tends to 1/2, which implies that the optimal strategy is as good as tossing a coin. In this case, the performance of the protocol is indistinguishable from that of an ideal one. This is known as "universally-composable security", and accounts for the strongest notion of cryptographic security (see [Can01] and [Mas09]).

Let us discuss the implications and limitations of our result. Note first that step 2 in the protocol involves a possible abortion (a similar step can also be found in Ref. [CR12b]). Hence, only those ϵ-sources with a non-negligible probability of passing step 2 can be amplified by our protocol. The abortion step can be relaxed by choosing a larger value of the constant c used for rejection. Yet, in principle, it could possibly exclude some of the ϵ-sources defined in (4.2.2). Notice, however, that demanding that step 2 is satisfied with non-negligible probability is just a restriction on the statistics seen by the honest parties $P(x_1, \ldots, x_n)$ and does not imply any restriction on the value of ϵ in $\epsilon \leq P(x_1, \ldots, x_n|e) \leq (1-\epsilon)$, which can be arbitrarily small. Also, we identify at least two reasons why sources that fulfil step 2 with high probability are the most natural in the context of randomness amplification. First, from a cryptographic perspective, if the observed sequence $x_1, \ldots, x_n$ does not fulfill step 2, then the honest parties will abort any protocol, regardless of whether a condition similar to step 2 is included. The reason is that such sequence would be extremely atypical in a fair source $P(x_1, \ldots, x_n) = 1/2^n$ and thus the honest players will conclude that the source is intervened by a malicious party or seriously damaged. Moreover, as discussed in Appendix B, imposing that the source has unbiased statistics from the honest parties' point of view does not imply any restriction on Eve's predictability. Second, from a more fundamental viewpoint, the question of whether truly random events exist in Nature is interesting since the observable statistics of many physical processes look random, that is, they are such that $P(x_1, \ldots, x_n) = 1/2^n$. If every process in nature was such the observable statistics does not fulfill step 2, the problem of whether truly random processes exist would hardly have been considered relevant. Finally, note that possible sources outside this subclass do not compromise the security of the protocol, only its probability of being successfully implemented.

Under the conditions demanded in step 2, our protocol actually goes through for sources more general than those in (4.2.2). These are defined by the following

restrictions,

$$G(n,\epsilon) < P\left(x_1,\ldots,x_n|e\right) \leq F(n,\epsilon), \tag{4.0}$$

for any pair of functions $G(n,\epsilon), F(n,\epsilon)$ defining the lower and upper bounds to Eve's control on the bias of each bit fulfilling the conditions $G(n,\epsilon) > 0$ and $\lim_{n\to\infty} F(n,\epsilon) = 0$. In fact, this condition is sufficient for our amplification protocol to succeed, see also Appendix B.

To complete the argument we must mention that, according to quantum mechanics, given a source that passes step 2, we can in principle implement the protocol with success probability equal to one, $P(g = 1) = 1$. It can be immediately verified that the qubit measurements X or Y on the quantum state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00000\rangle + |11111\rangle)$, with $|0\rangle$ and $|1\rangle$ the eigenstates of Z, yield correlations that maximally violate the five-partite Mermin inequality in question. (In a realistic scenario the success probability $P(g = 1)$ might be lower than one, but our Theorem warrants that the protocol is still secure.)

We can now state the main result of this Chapter.

Main result: A perfect free random bit can be obtained from sources of arbitrarily weak randomness using non-local quantum correlations.

We would like to conclude this Section by explaining the main intuitions behind the proof of the previous theorem. As mentioned, the protocol builds on the 5-party Mermin inequality because it is the simplest GHZ paradox allowing some extracted bit randomness certification. The estimation part, given by step 4, is rather standard and inspired by estimation techniques introduced in [BHK05], which were also used in [CR12b] in the context of randomness amplification. The most subtle part is the distillation of the final bit in step 5. Naively, and leaving aside estimation issues, one could argue that it is nothing but a classical processing by means of the function f of the imperfect random bits obtained via the N_d quintuplets. But this seems in contradiction with the result by Santha and Vazirani proving that it is impossible to extract by classical means a perfect free random bit from imperfect ones [SV86]. This intuition is however misleading. Indeed, the Bell certification allows applying techniques similar to those obtained in Ref. [Mas09] in the context of privacy amplification against non-signaling eavesdroppers. There, it was shown how to amplify the privacy, that is the unpredictability, of one of the measurement outcomes of bipartite correlations violating a Bell inequality. The key point is that the amplification, or distillation, is attained in a deterministic manner. That is, contrary to standard approaches, the privacy amplification process described in [Mas09] does not consume any randomness. Clearly, these deterministic techniques are extremely convenient for the randomness amplification scenario. In fact, the distillation part in our protocol can be seen as the translation

of the privacy amplification techniques of Ref. [Mas09] to our more complex scenario, involving now 5-party non-local correlations and a function of three of the measurement outcomes.

4.3. Discussion

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

From a more fundamental perspective, our results imply that there exist experiments whose outcomes are fully unpredictable. The only two assumptions for this conclusion are the existence of events with an arbitrarily small but non-zero amount of randomness that pass step 2 of our protocol and the validity of the no-signaling principle. Dropping the first assumption would lead to super-determinism, or to accept that the only source of randomness in nature are those that do not pass step 2 of our protocol, and in particular, that do not look unbiased. On the other hand, dropping the second assumption would imply abandoning a local causal structure for events in space-time. However, this is one of the most fundamental notions of special relativity.

5. Certifying observed randomness to be fully intrinsic

As discussed in Chapter 2, any observed random process includes two qualitatively different forms of randomness: apparent randomness, which results both from ignorance or lack of control of degrees of freedom in the system, and intrinsic randomness, which is not ascribable to any such cause. While classical systems only possess the first kind of randomness, quantum systems may exhibit some intrinsic randomness. However, all finite quantum processes known so far, including those of the previous Chapter, possess a mixture of both forms of randomness.

In this Chapter we study whether there exist finite quantum processess in which all the observed randomness is fully intrinsic. We provide an affirmative answer to this question under minimal assumptions: the impossibility of instantenous signaling and the existence of an arbitrarily weak source of initial randomness. Our results imply that there are quantum predictions that cannot be completed already in finite scenarios. The results presented in this Chapter are based on the article [DdlTA14].

5.1. Introduction

Non-local correlations observed when measuring entangled particles allow one to assess the randomness of a process independent of the full quantum formalism. As we saw in Chapter 2, under only two assumptions, (i) the impossibility of instantaneous communication, - known as the no-signaling principle - and (ii) that the measurement settings in a Bell test can be chosen at random - known as freedom of choice - non-local quantum correlations necessarily imply intrinsic randomness [CW12]. This is because such correlations cannot be described as the probabilistic mixture of *deterministic* processes.

Up to now, in all Bell tests, the intrinsic randomness revealed by quantum non-locality (under said assumptions) is also mixed with apparent randomness, resulting from the non-completeness of quantum theory. In this Chapter, we ask the following fundamental question: is there any quantum process that is as intrinsically random as it is observed to be and hence cannot be better

predicted? We answer this question in the affirmative by providing a family of quantum processes whose intrinsic randomness can be computed analytically for arbitrary system sizes and also demonstrating that this is strictly equal to the observed randomness. This implies that those events cannot be further completed, in the sense that a theory giving better predictions for these events should be either signalling or have no freedom of choice.

Our results are related to recent attempts to prove the completeness of quantum physics. In [CR11], Colbeck and Renner claimed that no no-signalling theory can have a better predictive power than quantum theory. However, the proof, which is based on the quantum violation of the chained Bell inequality, only works in an asymptotic regime where the number of measurements by the observers tends to infinite. That is, any finite scenario in [CR11] is such that quantum predictions can be completed. Moreover, the proof assumes that the settings in the inequality can be chosen freely. This last assumption leaves a significant space for improvement from a logical perspective since the free process needed in the proof is already assumed to be complete. A similar reasoning can be applied to the regular Bell theorem where the assumption of availability of initial perfect randomnes is referred to as the free-will assumption [KPB06, Hal11, Hal10, BG10, KHS$^+$12]. A possible way to strengthen the results on the completeness of quantum predictions is hence to weaken the said assumption by considering protocols for randomness amplification [CR12b]. There, the intrinsic randomness of a quantum process could be proven using a source of imperfect randomness. In fact, the protocol for full randomness amplification we gave in the previous Chapter provides a Bell test in which a measured variable has an intrinsic randomness that tends to be equal to the observed randomness in the limit of an infinite number of parties. All these proofs, then, left open a fundamental question: could it be the case that for all finite scenarios there always exists a gap between the randomness that we observe and that we can certify? What if the completeness of quantum theory is only an asymptotic property and, therefore, quantum predictions can be completed in any finite setup?

Our work answers these questions in the negative by providing finite-size Bell setups in which observed and intrinsic randomness are *strictly* equal. In fact, already an extremely simple Bell scenario consisting of three observers performing two dichotomic measurements produces events that cannot be completed. Moreover, our proof works using *arbitrarily small* randomness for the choice of measurements. In this sense, our results provide the strongest proof of completeness of quantum predictions.

Preliminaries

Suppose that a Bell test is performed repeatedly among N parties and the resulting statistics is given by $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$, where $\mathbf{a} = (a_1, \ldots, a_N)$ and $\mathbf{x} = (x_1, \ldots, x_N)$ are the string of outcomes and measurement inputs of the parties involved. Let g be a function acting on the measurement results $\mathbf{a}$. As explained in Chapter 2, Section 2.3, there are different physically relevant notions of randomness. We will now recall those definitions while adapting them to our particular situation in which (i) we consider the randomness on functions of outcomes and not the outcomes themselves and (ii) we allow for correlations between the preparation e and the measurement setting $\mathbf{x}$.

First, the *observed randomness* of g for measurements $\mathbf{x}$ is the randomness computed directly from the statistics. Operationally, this may be defined as the optimal probability of guessing the outcome of g for input $\mathbf{x}$,

$$G_{\text{obs}}(g, \mathbf{x}, P_{\text{obs}}) = \max_{k \in \text{Im}(g)} \; P_{\text{obs}}(g(\mathbf{a}) = k|\mathbf{x}). \tag{5.0}$$

where $\text{Im}(g)$ is the image of function g.

Moving to the definition of the *intrinsic randomness*, we saw that one should consider all possible preparations of the observed statistics in terms of no-signalling probability distributions. In our context, a particular preparation reads

$$P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = \sum_e p(e|\mathbf{x}) P_e^{\text{ex}}(\mathbf{a}|\mathbf{x}) \tag{5.0}$$

where the P_e^{ex} are extremal points of the no-signaling set [BLM$^+$05]. The terms $p(e|\mathbf{x})$ may depend on $\mathbf{x}$, which accounts for possible correlations between the preparation e and the measurement settings $\mathbf{x}$, given that the choice of measurements are not assumed to be free. Hence, we define the intrinsic randomness of a function g by optimizing over all possible non-signalling preparations of P_{obs} so as to minimize the randomness of g. In other words,

$$G_{\text{int}}^{NS}(g, \mathbf{x}, P_{\text{obs}}) = \max_{p(e|\mathbf{x}), P_e^{\text{ex}} \in \mathcal{NS}} \sum_e p(e|\mathbf{x}) G_{\text{obs}}(g, \mathbf{x}, P_e^{\text{ex}})$$

subject to:

$$\sum_e p(e|\mathbf{x}) P_e^{\text{ex}}(\mathbf{a}|\mathbf{x}) = P_{\text{obs}}(\mathbf{a}|\mathbf{x})$$

$$p(\mathbf{x}|e) \geq \delta \quad \text{with} \quad \delta > 0; \; \forall \, \mathbf{x}, e$$

where $G_{\text{obs}}(g, \mathbf{x}, P_e^{\text{ex}}) = \max_k \; P_e^{\text{ex}}(g(\mathbf{a}) = k|\mathbf{x})$ is also the intrinsic randomness of P_e^{ex}, since intrinsic and observed randomness must coincide for extremal

points of the non-signalling set. Note that the condition $p(\mathbf{x}|e) \geq \delta > 0$ allows for an arbitrary (but not absolute) relaxation of the freedom of choice assumption by allowing for arbitrary (yet not complete) correlations between the preparation and the measurement settings. Physically, this condition ensures that all measurement combinations appear for all possible preparations e (See [TSS13] for the significance of this condition). Notice as well that this definition of intrinsic randomness is weaker than the notion given in Section 2.3 where the freedom of choice was explicitly assumed.

An example of a source of randomness fulfilling this condition is a source of partially random bits, in which the bits can be partially predicted. This source was called in the previous Chapter a Santha-Vazirani source [SV86]. Note however that our definition allows sources more general than Santha-Vazirani sources.

From a cryptographic point of view, the observed randomness is the one perceived by the parties performing the Bell test, whereas the intrinsic randomness is that perceived by a non-signalling eavesdropper possessing knowledge of the preparation of the observed correlations and with the ability to arbitrarily (yet not fully) bias the choice of the measurement settings.

In general, G_{obs} is strictly larger than G_{int} for two reasons. First, because the observed correlations might not be pure in the quantum set of correlations. That is, Eve may hold a quantum purification of the correlations we observe. A more important reason which subsumes the previous one is that, in this Chapter, G_{int} is calculated finding decompositions of the observed correlations in term of no-signalling correlations. Recall that the set of non-signalling correlations is larger than the quantum set. This implies that Eve can hold a no-signaling purification of the correlations we obseve even if those are extremal within the quantum set.

When expressed in terms of these quantities, the results in [BKP06, CR12a] provide a Bell test in which G_{int} approaches G_{obs} (and to $1/2$) in the limit of an infinite number of measurements and assuming free choices, that is, $p(\mathbf{x}|e)$ in (5.1) is independent of e. The results in [CR12b] allow some relaxation of this last condition. The results of the previous Chapter, published in [GMDLT$^+$13], arbitrarily relax the free-choice condition and give a Bell test in which G_{int} tends to G_{obs} (and both tend to $1/2$) in the limit of an infinite number of parties. In this Chapter we provide a significantly stronger proof, as we allow the same level of relaxation on free choices and provide Bell tests in which $G_{\text{int}} = G_{\text{obs}}$ for any number of parties. Moreover, a perfect random bit is obtained in the limit of an infinite number of parties.

5.2. Results

5.2.1. Definition of the scenario

Our scenario consists of N parties where each performs two measurements of two outcomes. In what follows, we adopt a spin-like notation and label the outputs by ± 1. Then, any non-signalling probability distribution can be written as (for simplicity we give the expression for three parties, but it easily generalizes to an arbitrary number)

$$\begin{aligned} &P(a_1, a_2, a_3|x_1, x_2, x_3) = \\ &\frac{1}{8}\Big(1 + a_1\langle A_1^{(x_1)}\rangle + a_2\langle A_2^{(x_2)}\rangle + a_3\langle A_3^{(x_3)}\rangle + \\ &a_1a_2\langle A_1^{(x_1)}A_2^{(x_2)}\rangle + a_1a_3\langle A_1^{(x_1)}A_3^{(x_3)}\rangle + \\ &a_2a_3\langle A_2^{(x_2)}A_3^{(x_3)}\rangle + a_1a_2a_3\langle A_1^{(x_1)}A_2^{(x_2)}A_3^{(x_3)}\rangle\Big), \end{aligned} \tag{5.-2}$$

where $A_i^{(x_i)}$ denotes the outputs of measurement x_i by each party i.

In this scenario, we will test a specific family of Bell Inequalities introduced by Mermin [Mer90] and hence known as Mermin inequalities. A member of this family of inequalities was also used in the previous Chapter to devise a full randomness amplification protocol (see equation (4.2.3)). Note however that here we use another notation in terms of correlators that is more convenient for the purposes of this Chapter. These inequalities can be described in a recursive way through the description of the function related to that inequality. The Mermin function reads

$$M_N = \frac{1}{2}M_{N-1}(A_N^{(0)} + A_N^{(1)}) + \frac{1}{2}M'_{N-1}(A_N^{(0)} - A_N^{(1)}), \tag{5.-2}$$

where $M_1 = A_1^{(0)}$ and the function M'_{N-1} is that obtained from M_{N-1} after swapping $A_i^{(0)} \leftrightarrow A_i^{(1)}$. It can be seen that the function M_N consists of the sum, up to some signs, of 2^N (2^{N-1}) products of local observables for even (odd) N. Local models are such that $M_N \leq 1$. The maximal non-signalling violation violation of the inequality is equal to 2^N (2^{N-1}) for even (odd) N, that is all the products of observables appearing in the inequality are equal to ± 1. We study probability distributions that give this maximal violation and focus our analysis on a function f that maps the N measurement results into one bit as follows:

$$f(\mathbf{a}) = \begin{cases} +1 & n_-(\mathbf{a}) = (4j+2); \quad \text{with } j \in \{0, 1, 2, \ldots\} \\ \\ -1 & \text{otherwise} \end{cases} \tag{5.-2}$$

where $n_-(\mathbf{a})$ denotes the number of results in $\mathbf{a}$ that are equal to -1.

Our goal in what follows is to quantify the intrinsic randomness of the bit defined by $f(\mathbf{a})$ for those distributions maximally violating the Mermin inequality for odd N. We first prove the following

Lemma 5.1. *Let $P_{\mathrm{M}}(\mathbf{a}|\mathbf{x})$ be an N-partite (odd N) non-signalling probability distribution maximally violating the corresponding Mermin inequality. Then, for the input $\mathbf{x_m} = (0, \ldots, 0, 1)$ appearing in the inequality*

$$P_{\mathrm{M}}(f(\mathbf{a}) = h_N|\mathbf{x_m}) \geq 1/2, \text{ with } h_N = \sqrt{2}\cos\left(\frac{\pi(N+4)}{4}\right). \tag{5.-2}$$

Note that, as N is odd, $h_N = \pm 1$. Operationally, the Lemma implies that, for all points maximally violating the Mermin inequality, the bit defined by f is biased towards the same value h_N. Since the proof of the Lemma for arbitrary odd N is convoluted, we give the explicit proof for $N = 3$ here, which already conveys the main ingredients of the general proof, and relegate the generalization to Appendix C

Proof of lemma 5.1 for three parties. With some abuse of notation, the tripartite Mermin inequality may be expressed as,

$$M_3 = \langle 001\rangle + \langle 010\rangle + \langle 100\rangle - \langle 111\rangle \leq 2, \tag{5.-2}$$

where $\langle x_1x_2x_3\rangle = \langle A_1^{(x1)}A_2^{(x2)}A_3^{(x3)}\rangle$ and similar for the other terms. The maximal non-signalling violation assigns $M_3 = 4$ which can only occur when the first three correlators in (5.2.1) take their maximum value of $+1$ and the last takes its minimum of -1.

Let us take the corresponding input combination appearing in the inequality (5.2.1), $\mathbf{x_m} = (0, 0, 1)$. Maximal violation of M_3 imposes the following conditions:

1. $\langle 001\rangle = 1$. This further implies $\langle 0\rangle_1 = \langle 01\rangle_{23}$, $\langle 0\rangle_2 = \langle 01\rangle_{13}$ and $\langle 1\rangle_3 = \langle 00\rangle_{12}$.

2. $\langle 010\rangle = 1$ implying $\langle 0\rangle_1 = \langle 10\rangle_{23}$, $\langle 1\rangle_2 = \langle 00\rangle_{13}$ and $\langle 0\rangle_3 = \langle 01\rangle_{12}$.

3. $\langle 100\rangle = 1$ implying $\langle 1\rangle_1 = \langle 00\rangle_{23}$, $\langle 0\rangle_2 = \langle 10\rangle_{13}$ and $\langle 0\rangle_3 = \langle 10\rangle_{12}$.

4. $\langle 111\rangle = -1$ implying $\langle 1\rangle_1 = -\langle 11\rangle_{23}$, $\langle 1\rangle_2 = -\langle 11\rangle_{13}$ and $\langle 1\rangle_3 = -\langle 11\rangle_{12}$

Imposing these relations on (5.2.1) for input $\mathbf{x_m} = (0, 0, 1)$ one gets

$$\begin{aligned} P_{\mathrm{M}}(a_1, a_2, a_3|0, 0, 1) = \frac{1}{8}\,(1 + a_1a_2a_3 + (a_1 + a_2a_3)\langle 0\rangle_1 + \\ (a_2 + a_1a_3)\langle 0\rangle_2 + (a_3 + a_1a_2)\langle 1\rangle_3) \end{aligned} \tag{5.-2}$$

Using all these constraints and the definition of the function (5.2.1), Eq. (5.1) can be expressed as

$$\begin{aligned} &P_M(f(\mathbf{a}) = +1|\mathbf{x_m}) \\ &= P_{\mathrm{M}}(1, -1, -1|\mathbf{x_m}) + P_{\mathrm{M}}(-1, 1, -1|\mathbf{x_m}) \\ &+ P_{\mathrm{M}}(-1, -1, 1|\mathbf{x_m}) \\ &= \frac{1}{4}(3 - \langle 0\rangle_1 - \langle 0\rangle_2 - \langle 1\rangle_3) \end{aligned} \tag{5.-1}$$

Proving that $P(f(\mathbf{a}) = +1|\mathbf{x_m}) \geq 1/2$ then amounts to showing that $\langle 0\rangle_1 + \langle 0\rangle_2 + \langle 1\rangle_3 \leq 1$. This form is very convenient since it reminds one of a positivity condition of probabilities.

We then consider the input combination $\overline{\mathbf{x}}_\mathbf{m}$ such that all the bits in $\overline{\mathbf{x}}_\mathbf{m}$ are different from those in $\mathbf{x_m}$. We call this the swapped input, which in the previous case is $\overline{\mathbf{x}}_\mathbf{m} = (1, 1, 0)$. Note that this is *not* an input appearing in the Mermin inequality. However, using the previous constraints derived for distributions P_{M} maximally violating the inequality, one has

$$\begin{aligned} &P_{\mathrm{M}}(a_1, a_2, a_3|1, 1, 0) \\ =&\frac{1}{8}(1 + a_1\langle 1\rangle_1 + a_2\langle 1\rangle_2 + a_3\langle 0\rangle_3 + a_1a_2\langle 11\rangle_{12} \\ &+ a_1a_3\langle 10\rangle_{13} + a_2a_3\langle 10\rangle_{23} + a_1a_2a_3\langle 110\rangle_{123}) \\ =&\frac{1}{8}(1 + a_1\langle 1\rangle_1 + a_2\langle 1\rangle_2 + a_3\langle 0\rangle_3 - a_1a_2\langle 1\rangle_3 \\ &+ a_1a_3\langle 0\rangle_2 + a_2a_3\langle 0\rangle_1 + a_1a_2a_3\langle 110\rangle_{123}), \end{aligned} \tag{5.-1}$$

where the second equality results from the relations $\langle 11\rangle_{12} = -\langle 1\rangle_3$, $\langle 10\rangle_{13} = \langle 0\rangle_2$ and $\langle 10\rangle_{23} = \langle 0\rangle_1$.

It can be easily verified that summing the two positivity conditions $P_{\mathrm{M}}(1, 1, -1|\overline{\mathbf{x}}_\mathbf{m}) \geq 0$ and $P_{\mathrm{M}}(-1, -1, 1|\overline{\mathbf{x}}_\mathbf{m}) \geq 0$ gives the result we seek, namely $1 - \langle 0\rangle_1 - \langle 0\rangle_2 - \langle 1\rangle_3 \geq 0$, which completes the proof. □

5.2.2. Fully intrinsic observed randomness in finite-size systems

Using the previous Lemma, it is rather easy to prove our main theorem

Theorem 5.2 (Fully intrinsic randomness in finite-size systems)**.** *Let $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ be an N-partite (odd N) non-signalling probability distribution maximally violating the corresponding Mermin inequality. Then the intrinsic and the observed randomness of the function f are equal for the input $\mathbf{x_m}$ appearing in the Mermin inequality:*

$$G_{\text{int}}(f, \mathbf{x_m}, P_{\text{obs}}) = G_{\text{obs}}(f, \mathbf{x_m}, P_{\text{obs}})$$

where

$$G_{\text{obs}}(f, \mathbf{x_m}, P_{\text{obs}}) \quad = \quad \max_{k\in\{+1,-1\}} P_{\text{obs}}(f(\mathbf{a}) = k|\mathbf{x_m})$$

Proof of Theorem 5.2. Since P_{obs} maximally and algebraically violates the Mermin inequality, all the extremal distributions P_e^{ex} appearing in its decomposition must also necessarily lead to the maximal violation of the Mermin inequality (see Appendix C, Section C.2 for details). Hence, the randomness of f in these distributions as well satisfies Eqn. (5.1) of Lemma 5.1. Using this, we find,

$$\begin{aligned} G_{\text{obs}}(f, \mathbf{x_m}, P_e^{\text{ex}}) &= \max_{k\in\{+1,-1\}} P_e^{\text{ex}}(f(\mathbf{a}) = k|\mathbf{x_m}) \\ &= |P_e^{\text{ex}}(f(\mathbf{a}) = h_N|\mathbf{x_m}) - 1/2| + 1/2 \\ &= P_e^{\text{ex}}(f(\mathbf{a}) = h_N|\mathbf{x_m}), \end{aligned}$$

for every e. Therefore,

$$\begin{aligned} &G_{\text{int}}(f, \mathbf{x_m}, P_{\text{obs}}) \\ &\quad = \max_{\{p(e|\mathbf{x_m}), P_e^{\text{ex}}\}} \sum_e p(e|\mathbf{x_m}) G_{\text{obs}}(f, \mathbf{x_m}, P_e^{\text{ex}}) \\ &\quad = \max_{\{p(e|\mathbf{x_m}), P_e^{\text{ex}}\}} \sum_e p(e|\mathbf{x_m}) P_e^{\text{ex}}(f(\mathbf{a}) = h_N|\mathbf{x_m}) \\ &\quad = P_{\text{obs}}(f(\mathbf{a}) = h_N|\mathbf{x_m}), \end{aligned}$$

Likewise, the last equality follows from the constraint $\sum_e p(e|\mathbf{x}) P_e(\mathbf{a}|\mathbf{x}) = P_{\text{obs}}(\mathbf{a}|\mathbf{x})$. On the other hand the observed randomness for f is, $G_{\text{obs}}(f, \mathbf{x_m}, P_{\text{obs}}) = P_{\text{obs}}(f(\mathbf{a}) = h_N|\mathbf{x_m})$. □ □

5.2.3. Fully intrinsic observed random bit

The previous technical results are valid for any non-signalling distribution maximally violating the Mermin inequality. For odd N this maximal violation can

be attained by a unique quantum distribution, denoted by $P_{\text{ghz}}(\mathbf{a}|\mathbf{x})$, resulting from measurements on a GHZ state. When applying Theorem 5.2 to this distribution, one gets

Main result: Let $P_{\text{ghz}}(\mathbf{a}|\mathbf{x})$ be the N-partite (odd N) quantum probability probability distribution attaining the maximal violation of the Mermin inequality. The intrinsic and observed randomness of f for the Mermin input $\mathbf{x_m}$ satisfy

$$G_{\text{int/obs}}(f, \mathbf{x_m}, P_{\text{ghz}}) = \frac{1}{2} + \frac{1}{2^{(N+1)/2}} \tag{5.-8}$$

This follows straightforwardly from Theorem 5.2, since $P_{\text{ghz}}(\mathbf{a}|\mathbf{x}) = 1/2^{N-1}$ for outcomes $\mathbf{a}$ with an even number of results equal to -1 and for those measurements appearing in the Mermin inequality.

It is important to remark that $f(\mathbf{a})$ approaches a perfect random bit exponentially with the number of parties. In fact, this bit defines a process in which full randomness amplification takes place. Yet, it is not a complete protocol as, contrary to Chapter 4, no estimation part is provided.

5.3. Discussion

We have identified the first family of quantum processes whose observed randomness can be proven to be fully intrinsic. In other words, for the considered processes, quantum theory gives predictions as accurate as any no-signalling theory, possibly supra-quantum, can give and hence admit no further completion. Our results hold under the minimal assumptions: the validity of the no-signaling principle and an arbitrary (but not complete) relaxation of the freedom of choice. The latter is subtle and much attention in recent years has focused on relaxing it in Bell tests [KPB06, Hal11, Hal10, BG10, KHS$^+$12].

Our work raises several questions. Our main motivation here has been to understand the ultimate limits on the completeness of quantum theory for finite tests and, thus, we have worked in a noise-less regime. It is interesting to consider how would our results have to be modified to encompass scenarios including noise and hence amenable to experiments. The presence of noise modifies our results from two different viewpoints.

First, noise is due to lack of control of the setup and, thus, a source of apparent randomness, which immediately implies a gap between intrinsic and observed randomness. Second, in a noisy situation, it is impossible to arbitrarily relax the freedom of choice assumption, quantified by δ in Eq. (5.1). In fact, there is a tradeoff between the amount of relaxation of this condition and the

violation needed to certify the presence of any intrinsic randomness. The reason is that, for a sufficiently small value of δ, any correlations not attaining the maximal non-signalling violation of a Bell inequality can be reproduced using purely deterministic local strategies. For an explicit example of this tradeoff recall our treatment of the CHSH inequality in the previous Chapter, Section 4.2.3.

It thus seems natural, in a practical context, to extend the definition of intrinsic randomness by considering bounded relaxations of the freedom of choice assumption and non-maximal violations of Bell inequalities. These investigations could lead to stronger experimental tests on the completeness of quantum predictions, given that they would rely on significantly more relaxed assumptions than any other quantum experiment performed to date.

From a purely theoretical perspective, our results certify a maximum of one bit of randomness for any system size. It would be interesting to extend these analytical results to certify randomness that scales with the number of parties. This could for instance be accomplished with functions of increasing outcomes.

In a related context, it would also be interesting to explore whether similar results are possible in a bipartite scenario or, on the contrary, whether an asymptotic number of parties is necessary for full randomness amplification. We will discuss the recent advances on those questions in the Conclusions Chapter.

6. Nonlocality of a theory and its randomness capabilities

Correlations that violate a Bell Inequality are said to be nonlocal, i.e. they do not admit a local and deterministic explanation. Great effort has been devoted to study how the amount of nonlocality (as measured by a Bell inequality violation) serves to quantify the amount of randomness present in observed correlations. In particular, Chapters 4 and 5 contribute to this research program.

In this Chapter, however, we reverse this research program and ask what do the randomness certification capabilities of a theory tell us about the nonlocality of that theory. We find that maximal randomness certification cannot occur in maximally nonlocal theories. We go on and show that quantum theory, in contrast, permits certification of maximal randomness in all dichotomic scenarios. We hence pose the question of whether quantum theory is optimal for randomness. We answer this question in the negative by identifying a larger-than-quantum set of correlations capable of this feat. The results of this Chapter are based on the article [dlTHD$^+$14].

6.1. Introduction

In a device-independent protocol, no assumption is made about the inner-workings of the devices used and are thus regarded as black boxes. Recall from Chapter 2, Section 2.2.2 that there is however a crucial assumption to every protocol and that is the assumption of the background theory dictating the devices' behaviour, e.g. whether the devices are quantum mechanical [ABG$^+$07, RUV13]), or just compatible with no-signaling principle [BHK05]). In this Chapter we will elaborate on the implications for randomness certification of assuming these different theories.

The assumption about the background theory is *vital* given that quantum mechanics is not the most nonlocal theory respecting the no-signaling principle [PR94] and therefore capable of producing intrinsic randomness. Theories allowing for all nonlocal correlations *only* restricted by the no-signaling principle are termed "maximally nonlocal" throughout this thesis since they produce the most amount of nonlocality that a non-signalling theory can produce. Given

the eminent role of nonlocality for randomness certification, the first intuition is to expect maximally nonlocal theories to have more powerful randomness certification capabilities than theories with limited nonlocality, such as quantum mechanics. Indeed, there are occasions where maximally nonlocal theories can certify randomness and quantum mechanics cannot even certify any randomness at all [ABB$^+$10]. On the other hand, for the Clauser-Horne-Shimony-Holt (CHSH) inequality [CHSH69], we can certify more randomness assuming only quantum mechanics (however not the maximal amount possible) rather than allowing maximally nonlocal correlations [AMP12a].

The main goal of this Chapter is to understand the relationship between the nonlocality and randomness of a theory and, in particular, what the randomness capabilities of a theory tell us about the nonlocality allowed within that theory.

The first result is to show that the previous intuition is wrong: were the set of achievable physical correlations not more restricted than what the no-signaling principle allows, maximal randomness could not be certified in any possible scenario irrespective of the number of parties, measurements or outcomes, i.e. maximally nonlocal theories cannot be maximally random.

Secondly, we focus on quantum theory and provide, in contrast, scenarios with an arbitrary number of parties where maximal randomness *can* be certified. This should be compared with other works that showed that if maximally nonlocal theories were permitted in Nature we would have unimaginable computational and communicating power [PPK$^+$09, BBL$^+$06]. Here, being in a maximally nonlocal world *limits* our information processing capabilities. This observation leads us to ask if the nonlocality of quantum theory is in some sense optimal for randomness certification. That is, is quantum theory the most nonlocal theory capable of certifying maximal randomness? Our final result answers this question in the negative: we identify a set of correlations larger than the quantum set that also permits the certification of maximal randomness.

6.2. Results

The results stated in this Section will use the definitions of observed and intrinsic randomness we gave in Chapter 2, Section 2.3.

6.2.1. Maximally nonlocal correlations

Recall from Chapter 2, Section 2.2.2 that the set $\mathcal{NS}$ of maximally nonlocal correlations is the set of multipartite correlations solely restricted by the no-signaling principle. Let us state the first result concerning this set of correlations.

Result 1: *Maximally nonlocal theories can never be maximally random.*

Were the physically achievable correlations solely restricted by the no-signaling principle, the maximum amount of certifiable randomness in an arbitrary Bell scenario (N, M, d) would be bounded through the intrinsic randomness by

$$G_{\text{int}}^{NS}(\mathbf{x}_0, P_{\text{obs}}) \geq \frac{1}{d^N - (d-1)^N}, \tag{6.0}$$

for any probability distribution $P_{\text{obs}} \in \mathcal{NS}$ and all inputs $\mathbf{x}_0$.

To prove this result we only need to consider the randomness of the extreme points P_e^{ex} of $\mathcal{NS}$ as indicated by (2.7). Our proof is based on the simple observation that if for a particular $\mathbf{x}_0$ of correlations $P(\mathbf{a}|\mathbf{x})$, n values are equal to zero then $\max_{\mathbf{a}} P(\mathbf{a}|\mathbf{x}_0) \geq \frac{1}{d^N - n}$. Result 1 follows from the following Theorem which provides an upper bound on the number of non-zero entries in extreme non-signaling correlations.

Theorem 6.1. *Let $P(\boldsymbol{a}|\boldsymbol{x})$ be an extreme probability distribution in the set $\mathcal{NS}$ in an arbitrary Bell test scenario (N, M, d). For a given combination of settings $\boldsymbol{x}_0$, denote by $n(\boldsymbol{x}_0)$ the number of probabilities $P(\boldsymbol{a}|\boldsymbol{x}_0)$ that are equal to zero and define $n = \min_{\boldsymbol{x}_0} n(\boldsymbol{x}_0)$. Then, $n \geq (d-1)^N$.*

Proof. The proof of the result follows from a relatively simple counting argument. First we introduce some useful notation to describe the marginals of a probability distribution. If we have a distribution P with elements $P(\mathbf{a}|\mathbf{x})$ and we have a set $\mathcal{J} \subseteq \{1, 2, ..., N\}$ of the N parties then the probability distribution only over these parties in $\mathcal{J}$ is $P(\mathbf{a}^{\mathcal{J}}|\mathbf{x}^{\mathcal{J}}) = \sum_{a_j | j \notin \mathcal{J}}^{N} P(\mathbf{a}|\mathbf{x})$ where $\mathbf{a}^{\mathcal{J}}$ and $\mathbf{x}^{\mathcal{J}}$ are $\mathbf{a}$ and $\mathbf{x}$ consisting only of elements a_j and x_j respectively for all $j \in \mathcal{J}$. Following a simple generalization of Ref. [CG04], a probability distribution (for any input $\mathbf{x}_0$) satisfying the no-signalling principle can be parametrized by $P(\mathbf{a}^{\mathcal{J}}|\mathbf{x}^{\mathcal{J}})$ for all possible sets $\mathcal{J}$. What is more, due to normalization we only consider $(d-1)$ outputs for each party in all of these distributions. Therefore the probability $P(\mathbf{a}|\mathbf{x})$ is a function of $P(\mathbf{a}^{\mathcal{J}}|\mathbf{x}^{\mathcal{J}})$ for all $\mathcal{J}$ but the elements a_j of $\mathbf{a}^{\mathcal{J}}$ only range over $(d-1)$ values. Apart from when $\mathcal{J}$ contains all N parties, every other marginal probability $P(\mathbf{a}^{\mathcal{J}}|\mathbf{x}^{\mathcal{J}})$ will result from another probability distribution $P(\mathbf{a}|\mathbf{x}')$ for $\mathbf{x}' \neq \mathbf{x}_0$ by summing over outputs of the appropriate parties. Therefore the values of these marginals are fixed by probabilities for inputs $\mathbf{x}' \neq \mathbf{x}_0$ and the only free parameters defining $P(\mathbf{a}|\mathbf{x}_0)$ are the $(d-1)^N$ probabilities when $\mathcal{J}$ contains all N parties.

Clearly, the space of this d^N-outcome probability distributions $P(\mathbf{a}|\mathbf{x}_0)$ is convex. Moreover, if $P(\mathbf{a}|\mathbf{x}_0)$ is not an extreme point in this space, neither

are the original correlations $P(\mathbf{a}|\mathbf{x})$ in the original non-signalling space. As mentioned, when restricted to the specific setting $\mathbf{x}_0$, there are $(d-1)^N$ free parameters. Now, the hyperplanes defining this convex space correspond to the positivity constraints defined by the d^N probabilities $P(\mathbf{a}|\mathbf{x}_0)$. An extreme point in this space of dimension $(d-1)^N$ should then be defined by the intersection of $(d-1)^N$ hyperplanes. This implies that a necessary condition for the correlations $P(\mathbf{a}|\mathbf{x})$ to be extreme is that at least $(d-1)^N$ probabilities $P(\mathbf{a}|\mathbf{x}_0)$ are zero for each value of $\mathbf{x}_0$. This completes the proof. □

It is worth mentioning two facts. First, this result indicates an important limitation on maximally nonlocal theories. In fact, the gap between the ideal maximal randomness and that achievable in maximally nonlocal theories is unbounded. Second, the derived bound is, in general, not tight. For instance, all extreme non-signaling correlations in Bell test scenarios $(2, M, 2)$ were obtained in [JM05, BP05] and in this case $G^{NS}_{\text{int}}(\mathbf{x}_0, P_{\text{obs}}) \geq 1/2$ whereas our bound gives $1/3$. Interestingly, the same difference appears in the $(3, 2, 2)$ scenario: looking at all the extreme points, classified in [PBS11], the maximal randomness is equal to $1/6$, while our bound predicts $1/7$. However, in the asymptotic limit of $d \to \infty$ our bound gives $\frac{1}{O(d^{N-1})}$, which can be shown to be tight by comparing it with the results in [AGCA12]. We now move to randomness certification in quantum theory.

6.2.2. Quantum correlations

Recall from the Preliminaries Chapter that a probability distribution $P_{\text{obs}} \in \mathcal{Q}$ belongs to the quantum set of correlations if it can be written as $P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = \text{tr}(\rho \bigotimes_{j=1}^{N} O_{a_j}^{x_j})$.

Characterizing the set of correlations achievable in this way is a great open problem in quantum information theory. Therefore, in what follows, rather than solving exactly the optimization problem (2.7), we consider a relaxation that provides a lower bound to the intrinsic randomness. Instead of considering all convex combinations of extreme points of $\mathcal{Q}$ that reproduce the observed statistics, we ask for convex combination of extreme points that give an observed violation of a Bell inequality. Given that a Bell inequality is just a linear combination of probabilities $P(\mathbf{a}|\mathbf{x})$ over all inputs $\mathbf{a}$ and outputs $\mathbf{x}$, let us define the following inner product between correlations P_{obs} and Bell inequality B that computes the Bell violation $B \cdot P_{\text{obs}} \equiv \sum_{\mathbf{a},\mathbf{x}} \beta_{\mathbf{a},\mathbf{x}} P_{\text{obs}}(\mathbf{a}|\mathbf{x}) = q_{\text{obs}}$, where the real coefficients $\beta_{\mathbf{a},\mathbf{x}}$ define the Bell inequality B.

Computing a lower bound to the intrinsic randomness $G^{Q}_{\text{int}}(\mathbf{x}_0, P_{\text{obs}})$, certified this time by an observed violation of a Bell inequality, then amounts to solving

the following optimization problem, a relaxation of (2.7):

$$G^{Q}_{\text{int}}(\mathbf{x}_0, P_{\text{obs}}) \leq \max_{p(e), P^{\text{ex}}_e \in \mathcal{Q}} \sum_e p(e) G_{\text{obs}}(\mathbf{x}_0, P^{\text{ex}}_e)$$

subject to:

$$\sum_e p(e)(B \cdot P^{\text{ex}}_e) = q_{\text{obs}}, P^{\text{ex}}_e \in \mathcal{Q}.$$

Since we are interested in the maximal amount of randomness allowed by quantum mechanics, we will restrict our study to maximal quantum violation of a Bell Inequality $q_{\text{obs}} \equiv q_{\max}$. In [DPA13] a method was provided to detect when the maximal quantum violation of a Bell inequality certifies that the outputs are maximally random. The method has the advantage that it can be easily applied, but unfortunately it only works under the assumption that the maximal quantum violation of the inequality is unique. The uniqueness of the maximal quantum violation is in general hard to prove. However, in what follows, we consider Bell inequalities for which the uniqueness of the maximal violation can be proven using the results of Refs. [FFW11, MS13]. This then allows us to apply the simple method in [DPA13] and prove the following result.

Result 2: *Quantum theory is maximally random in all dichotomic scenarios.*
Assuming the set of physically achievable correlations to be the quantum set, the maximum amount of certifiable randomness in the family of Bell test scenarios $(N, M, 2)$ is maximal: $G^{Q}_{\text{int}}(\mathbf{x}_0, P_{\text{obs}}) = \frac{1}{2^N}$.

We prove this result in Appendix D by generalizing the results of [DPA13] to all N via a Bell inequality introduced in [HCLB11]. We actually prove Result 2 for the $(N, 2, 2)$ scenario but this trivially applies to the $(N, M, 2)$ since we can always ignore $(M - 2)$ of the inputs for each party. While our proof does not apply to the case of two parties, it has been shown in [AMP12a] that for the $(2, 2, 2)$ scenario an amount of randomness arbitrarily close to the maximum of 2 random bits can be certified in some limit. Additionally, it has been shown analitically that exactly 2 bits of maximal randomness can be attained in the $(2, 3, 2)$ scenario [Sca14]. All of this serves to show that quantum correlations certify maximal randomness even if maximally nonlocal theories can never do this.

We have shown the difference for randomness certification of two sets of correlations; the maximally nonlocal set and the quantum set. A natural question is whether this contrast highlights the *uniqueness* of quantum correlations: not only can we certify the generation of randomness (something impossible in classical physics) but we can certify *maximal* randomness (something impossible in

maximally nonlocal theories). Perhaps the certification of maximal randomness could be an *information principle* that allows us to recover quantum mechanical correlations from the set of all correlations that satisfy no-signaling. Other examples of information principles include Information Causality [PPK$^+$09], Non-trivial Communication Complexity [BBL$^+$06] and Local Orthogonality [FSA$^+$13]. Is it possible to recover quantum correlations utilising the principle that a theory should certify maximal randomness? In other words, are there other sets of correlations that allow for maximal randomness certification? We now address this question.

6.2.3. Supra-quantum correlations

Navascués, Pironio and Acín introduced a means to approximate the set of quantum correlations which was an infinite hierarchy of semi-definite programs [NPA07]. This hierarchy has an infinite number of levels where each level defines a set of correlations defined in terms of a positive semi-definite matrix. For example, the first non-trivial level of this hierarchy is $\mathcal{Q}^1$ and this set is provably larger than the set of quantum correlations $\mathcal{Q}$. Already in the work of Pironio et al in Ref. [PAM$^+$10c] these first few levels in the hierarchy were used to lower bound the amount of randomness certified for quantum correlations. Since the first few levels of the hierarchy produce supra-quantum correlations, we could expect these sets of correlations not to give maximal randomness. Our third main result that this intuition is incorrect by introducing a set of correlations that can produce maximal randomness. In Appendix D, we introduce a modification to the set $\mathcal{Q}^1$ in the tripartite setting called $\mathcal{Q}^{1+ABC}$ that is strictly larger than the quantum set and allows for maximal randomness certification. This represents the third main result of this work.

Result 3: *There exist post-quantum theories that can also certify maximal randomness.*

Were the physically achievable correlations those of the strictly larger than quantum set $\mathcal{Q}^{1+ABC}$, maximal randomness could also be certified in the Bell test scenario $(3, M, 2)$ i.e. $G_{\text{int}}^{\mathcal{Q}^{1+ABC}}(\mathbf{x}_0, P_{\text{obs}}) = \frac{1}{8}$.

The proof of this result is presented in Appendix D. The crucial element in this proof is showing that there is only one probability distribution in the set $\mathcal{Q}^{1+ABC}$ that maximally violates the Mermin inequality [Mer90] allowing us to use the results in Ref. [DPA13].

At first, this result may seem disappointing but there are other examples of limitations to recovering quantum correlations from information principles.

For example it is known that we need truly multipartite principles [GWAN12]. It has also been shown that other information principles will never recover quantum mechanical correlations [NGHA14] and our work fits squarely within this foundational research program.

6.3. Discussion

In this Chapter we have shown that correlations in maximally nonlocal theories and quantum theory have drastically different consequences for randomness certification. Therefore, if we assume Nature does not abide by a nonlocality-restricted theory such as quantum theory it could severely limit its randomness capabilities. In turn, the bounded amount of nonlocality in quantum theory is enough for maximal randomness certification. The proof of our results also gives an explanation for this a priori counter-intuitive effect. The maximal certifiable randomness of a theory is determined by its extreme correlations. In a maximally nonlocal theory, extreme points display correlations for any combination of inputs: as proven, some outcomes across all parties are necessarily excluded for any combination of inputs for correlations to be extreme. But, if we want to certify maximal randomness then every outcome needs to occur. Quantum extreme correlations, on the other hand, being a mixture of maximally nonlocal extreme correlations, permit maximal randomness.

These results are not only of foundational interest but have application in randomness extraction, certification and amplification. For example, in Ref. [PAM$^+$10c] a lower bound on certifiable randomness was obtained using only the no-signaling principle, and this bound has found applications in other protocols (e.g. Ref. [VV12b]). Not only do we show that maximal randomness is impossible for maximally nonlocal theories but we also give a bound on this randomness. Hopefully this bound will be useful in the design and analysis of future protocols. An interesting follow-up question is to determine the *exact* maximum randomness allowed just by the no-signalling principle, a fundamental number providing a quantitative link between randomness and no-signalling.

In contrast, we have shown that quantum theory allows for maximal randomness certification where parties perform dichotomic measurements. However, we conjecture that quantum correlations can produce certifiable maximal randomness for all scenarios. To finish, we have also shown that maximal randomness certification is a property shared by other supra-quantum theories. We have given an example of a more nonlocal theory with the same ability in the case of three parties performing dichotomic measurements. This last result indicates that quantum theory is not so special from an information theoretic perspec-

tive (cf. Ref. [NGHA14]). Also the set of quantum correlations is notoriously difficult to define whereas the set of maximally nonlocal correlations is defined by linear inequalities therefore there is an apparent trade off in the complexity of the set of correlations with the randomness that can be obtained from it. The fact that there exists a set of correlations that has a relatively simple description but facilitates maximal randomness certification provides a "third way" for the design and analysis of future protocols.

7. Quantum correlations under local dimension constraints

Progress on semi-device-independent quantum information protocols depends crucially on our ability to bound the strength of the nonlocal correlations achievable with finite dimensional quantum resources.

In this Chapter we will show a method to characterize quantum nonlocality under local dimension constraints via a complete hierarchy of semidefinite programming relaxations. We will also show some applications of our method to semi-device-independent problems such as certifying high-dimensional entanglement in a device-independent way or detecting entangled measurements. The results presented in this Chapter are based on the article [NdlTV14].

7.1. Introduction

The realization by John Bell in his 1964 seminal paper [Bel64] that the correlations arising from measuring separated quantum systems so that the measurements define space-like separated events (quantum correlations) can be nonlocal represents one of the most outstanding discoveries of modern physics. The signature of nonlocality, the violation of a Bell inequality, has been extensively verified experimentally and stands as a well established experimental fact [PCL+12, BCP+14].

Besides its foundational interest, quantum nonlocality is instrumental in the emergent field of device-independent quantum information processing, whose objective is to infer properties of the underlying state and measurements without assuming any a priori knowledge of the inner working of the devices used. Quantum key distribution [ABG+07, PAB+09, MPA11, PMLA13], randomness generation [Col07, PAM+10c, GMDLT+13] and genuine multipartite entanglement certification [Hal10, BBS+13] are celebrated instances of information-theoretic tasks which can be implemented in a black box scenario. The characterization of quantum nonlocality provided by the Navascués-Pironio-Acín (NPA) hierarchy [NPA07] played a pivotal role in assessing the security of many of such protocols.

In the last years, it has been pointed out that in many physical situations, e.g., in ion-trap experiments, there is a bound or a promise on the dimensionality of the system under study. Exploiting this promise has led to *semi-device-independent* bounds on entanglement [LVB11] and novel quantum key distribution [PB11] and randomness generation [LPY$^+$12] protocols more robust and efficient than their fully device-independent counterparts. This approach to quantum information science stems from prior research on dimension witnesses [BPA$^+$08, PV08, PGWP$^+$08, WCD08, BBT11], which are device-independent lower bounds on the Schmidt rank of the bipartite state giving rise to the observed correlations. Clearly, in order to certify the security of semi-device-independent communication protocols, or the existence of high-dimensional entanglement in a device-independent way, a characterization of quantum correlations under local dimension constraints is needed. In this respect, see-saw variational techniques have proven very useful to characterize such a set of correlations from the inside [WW01, PV10].

Characterizations from the outside -i.e., the characterization of *limits*- are, on the contrary, problematic. A brute-force approach, advocated in [EHGC04], is to reduce the computation of Tsirelson bounds to the minimization of a multivariate polynomial over a region defined by polynomial constraints and run the Lasserre-Parrilo hierarchy of semidefinite programming relaxations [Las01, Par00]. Unfortunately, the vast amount of free variables needed to model the simplest nonlocality scenarios makes this scheme intractable in normal computers. Another possibility is to make use of the interesting algorithm proposed by Moroder *et al.* [MBL$^+$13]. This method works by implementing a modified version of the NPA hierarchy with extra positivity constraints which effectively bound the negativity [VW02] of the underlying quantum state $|\psi\rangle$. By restricting the negativity to be below the value $1/2$, this tool was successfully employed in [MBL$^+$13] to derive the maximum violations of the I_{3322} and I_{2233} inequalities [Fro81, CGL$^+$02] attainable with qubit systems. In principle, this method can be improved by imposing that not only the state ψ has negativity smaller than $1/2$, but also suitable local postselections of the form $P_A P_B |\psi\rangle$, where P_A (P_B) denotes a polynomial of Alice's (Bob's) measurement operators. It is not clear, though, that even this modified scheme converges to the desired set of correlations: indeed, since Peres' conjecture was shown to be false [Per99, VB14], there could exist high dimensional states with positive partial transpose which nevertheless produce correlations impossible to reproduce with, say, two qubits.

In the present Chapter, we introduce practical numerical techniques for the full characterization of quantum correlations in scenarios where the local dimension of some parts of a multipartite quantum system are trusted to be bounded, while the local dimensions of the rest of the parties stay fully unconstrained.

By exploiting a previously unnoticed connection with the separability problem, we show how to use tools from entanglement detection to characterize the strength of bipartite quantum correlations under local dimension constraints via hierarchies of semidefinite programming relaxations. Combined with the formalism of moment matrices from the NPA hierarchy, the resulting method becomes capable to deal with multipartite scenarios where a subset of the N parties has access to infinite dimensional degrees of freedom. In both cases, the convergence of our sequence of relaxations to the appropriate set of correlations is rigorously proven.

The application of these techniques to several device-independent problems is also studied. We use our method to bound the maximal violation attainable via measurements on two-qubit states of a number of bipartite Bell inequalities. This question arises naturally in quantum information science [BPA$^+$08, PV08] and convex optimization theory [BdOFV10], where high performance algorithms to solve the problem are still missing. In addition, we use our tools to derive a tripartite Bell inequality that allows to certify, for the first time, three-dimensional tripartite entanglement in a device-independent way, thereby extending Huber & de Vicente's recent work on multidimensional entanglement [HdV13] to the black box realm. We conclude with a semi-device-independent application: in [VN11] it was proposed a scheme to certify entangling dichotomic measurements under the assumption that the probed states are pairs of independent qubits. We make use our new numerical tools to prove that such a scheme works, i.e., that the linear witness presented in [VN11] does actually discriminate separable from entangled measurement operators.

7.2. Results

In this section we will describe two methods to characterize quantum nonlocality.

The first method described in Section 7.2.1 allows one to find upper bounds to the maximal violation of a given Bell Inequality when the local dimension of both parts of a bipartite quantum system is known. It is divided in two *modules*. The first module introduces a relaxation to our problem via positive partial transposition constraints. The second module is a refinement of the first and uses techniques of entanglement detection. It entails in turn a hierarchy of relaxations converging to the wanted solution ie. ultimately producing a tight bound.

The second method described in Section 7.2.2 allows one to find upper bounds to the maximal violation of a given Bell Inequality but when the local dimen-

sion of *some* parts of a multipartite quantum system is known, while the other remain fully unconstrained. Notice that, from the Schmidt decomposition, optimizing over bipartite scenarios were both parties have dimension constraints (ie. their measurents act on $B(\mathbb{C}^d)$) is equivalent to the scenario in which only one party has the dimension constraint (say Alice's measurements act on $B(\mathbb{C}^d)$ while Bob's are unconstrained). As we will see in the Applications Section, this fact will make the second method cheaper than the first to deal with some complex bipartite scenarios.

7.2.1. Bipartite nonlocality in finite dimensions

Module 1

Consider two parties, Alice and Bob, interacting with a two-qubit system via measurement devices. The measurement devices allow them to implement m different dichotomic measurements. Denoting Alice's and Bob's measurement settings by x and y and their measurement outputs by a and b, we hence have that x, y range from 1 to m, while a, b can take values in $\{0, 1\}$.

Call $Q(\mathbb{C}^2)$ the set of local and nonlocal probability distributions $P(a, b|x, y)$ which Alice and Bob can generate with this setting. The physical significance of $Q(\mathbb{C}^2)$ is quite clear: if we have the promise that the form of Alice and Bob's state and measurement operators does not vary during the course of the experiment, we should expect to observe distributions in $Q(\mathbb{C}^2)$. That set is in general non convex [PV09a]. A more realistic model, though, would contemplate the possibility that each physical realization of the experiment could be different from the previous one. In this scenario, the quantum resources used in the preparation have dimension constraints but the shared randomness used in the preparation is for free.

We will thus be concerned with the problem of conducting linear optimizations over $Q(\mathbb{C}^2)$, or, equivalently, characterizing the convex hull of this set.

Since we are speaking about dichotomic measurements, the extreme points of $Q(\mathbb{C}^2)$ are generated by conducting local projective measurements over a two-qubit state. Let us for the moment restrict to extreme points where all such measurements are rank-one projectors (degenerate cases can be treated in a similar manner), i.e.,

$$P(a, b|x, y) = \mathrm{tr}\{\rho_{AB}(\Pi_a^x \otimes \overline{\Pi}_b^y)\}, \tag{7.0}$$

where $\Pi_a^x = a\mathbb{I}_2 + (-1)^a |u^x\rangle\langle u^x|$, $\overline{\Pi}_b^y = b\mathbb{I}_2 + (-1)^b |v^y\rangle\langle v^y|$, and $|u^x\rangle, |v^y\rangle \in \mathbb{C}^2$ are normalized vectors.

We will now show that there is an equivalent way of writing $P(a,b|x,y)$ which will turn out to be very useful. For that, we will map Alice and Bob's state and measurement operators to a $(2m+2)$-qubit state W acting on a Hilbert space $ABA_1\cdots A_mB_1\cdots B_m$. The two-qubit space AB will store the shared quantum state ρ_{AB}, while the Hilbert spaces $A_1\cdots A_m$ ($B_1\cdots B_m$) will hold Alice's (Bob's) measurement projectors $\{|u^x\rangle\langle u^x|\}_{x=1}^m$ ($\{|v^y\rangle\langle v^y|\}_{y=1}^m$). The state W is thus given by

$$W = \rho_{AB} \otimes \bigotimes_{x=1}^{m} |u^x\rangle\langle u^x| \otimes \bigotimes_{y=1}^{m} |v^x\rangle\langle v^x| \,. \tag{7.0}$$

We can now write

$$P(a,b|x,y) = \mathrm{tr}(W M_a^x \otimes N_b^y), \tag{7.0}$$

where

$$\begin{aligned} M_a^x &= (a\mathbb{I}_{AA_x} + (-1)^a V(A,A_x)) \otimes \mathbb{I}_{A_1\cdots A_{x-1}A_{x+1}\cdots A_m} \\ N_b^y &= (a\mathbb{I}_{BB_y} + (-1)^b V(B,B_y)) \otimes \mathbb{I}_{B_1\cdots B_{y-1}B_{y+1}\cdots B_m} \end{aligned} \tag{7.-1}$$

Here $V(C,D)$ denotes the SWAP operator between the Hilbert spaces C, D. The SWAP $V(C,D)$ operator is defined as $V = \sum_{i,j=0}^{1} |i\rangle_C |j\rangle_D \langle j|_C \langle i|_D$. Note that W is a normalized quantum state, fully separable with respect to the partition $AB|A_1|\cdots|A_m|B_1|\cdots|B_m|$.

Conversely, it is easy to see that the convex hull of the set of all distributions $P(a,b|x,y)$ achievable by conducting rank-one measurements over a two-qubit state is given by all $P(a,b|x,y) = \mathrm{tr}(W M_a^x \otimes N_b^y)$, with W fully separable.

Consequently, finding the maximal violation of any Bell inequality $I = \sum_{a,b,x,y} B_{ab}^{xy} P(a,b|x,y)$ in the above systems is equivalent to solve the problem

$$\begin{aligned} & \max \mathrm{tr}(W \cdot \sum_{a,b,x,y} B_{ab}^{xy} M_a^x \otimes N_b^y), \\ \text{s.t.} \quad & \mathrm{tr}(W) = 1, W \geq 0 \\ & W, \text{ separable.} \end{aligned} \tag{7.-2}$$

Unfortunately, optimizing linearly over the set of separable states is an NP-hard problem [Gur04, Gha10]. Consider then the corresponding Positive Partial Transpose (PPT) [Per96] relaxation

$$\begin{aligned} & \max \operatorname{tr}(W \cdot \sum_{a,b,x,y} B_{ab}^{xy} M_a^x \otimes N_b^y), \\ \text{s.t.} \quad & \operatorname{tr}(W) = 1, W \geq 0, \\ & W^{T_P} \geq 0, \text{ for all bipartitions } P, \end{aligned} \tag{7.-3}$$

where W^{T_P} denotes the partial transpose of matrix W with respect to the systems P. Note that this condition is a relaxation of the tensor product form of the separable state constraint of the previous problem.

The above problem can be cast as a semidefinite program, and its solution will provide an upper bound on the violation of the said inequality. Note also that we can fix $|u^m\rangle = |v^m\rangle = |0\rangle$, and so, by modifying appropriately the definition of the operators M_a^x, N_b^y, we 'only' need to optimize over a $2 + 2(m-1) = 2m$-qubit state.

Module 2

In general, though, we should expect this method not to return the exact solution. This leads us to consider tighter relaxations of the separability condition. Let us now review briefly the Doherty-Parrilo-Spedalieri (DPS) hierarchy [DPS05] of semidefinite programs to detect entanglement. Afterwards, we will see how to use it to characterize quantum correlations with dimension constraints.

The intuition behind the DPS method is the observation that any fully separable state

$$\rho_{1,2,\ldots} = \sum_k p_k \left|u_k^1\right\rangle\!\left\langle u_k^1\right| \otimes \left|u_k^2\right\rangle\!\left\langle u_k^2\right| \otimes \ldots \otimes \left|u_k^n\right\rangle\!\left\langle u_k^n\right| \tag{7.-3}$$

admits an N extension per site of the form

$$\sigma \equiv \sum_k p_k \left|u_k^1\right\rangle\!\left\langle u_k^1\right|^{\otimes N_k} \otimes \ldots \otimes \left|u_k^{n-1}\right\rangle\!\left\langle u_k^{n-1}\right|^{\otimes N_k} \otimes \left|u_k^n\right\rangle\!\left\langle u_k^n\right|. \tag{7.-3}$$

for any $N \geq 2$. Note that we are not extending the last subsystem. The new state σ has the following properties:

1. It lives in the Hilbert space $\bigotimes_{k=1}^{n-1} \mathcal{H}_{d_k}^{N_k} \otimes \mathcal{H}_n$, where $\mathcal{H}_d^N$ denotes the N-symmetric space of $\mathbb{C}^d$.
2. It satisfies $\operatorname{tr}_{1^{N_1-1}2^{N_2-1}\ldots}(\sigma) = \rho$.

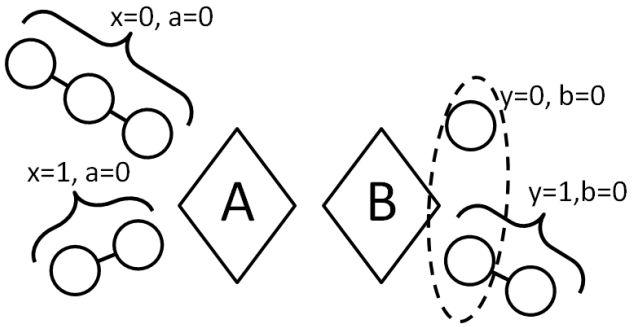

Figure 7.1.: Pictorial representation of the state constraints: the diamonds (heads) denote Alice and Bob state spaces; the circles represent rank-one projectors. N circles joined by a line (legs) must be understood as living in the symmetric space of N particles. In the figure, Alice's projectors $|u^1\rangle\langle u^1|$, $|u^2\rangle\langle u^2|$ are represented by legs of length 3, 2 respectively. Also, it has been imposed positivity under the partial transpose of one of the circles of Bob's first leg and Bob's second leg.

3. It is PPT with respect to all bipartitions.

A necessary condition for ρ to be fully separable is thus that a state with the above properties exists. It is easy to see that checking for the existence of such a state can be cast as a semidefinite program. Furthermore, in [DPS05] it is proven that the resulting entanglement criteria is complete, even when the last condition is omitted.

After reviewing the DPS method to detect entanglement (7.-2), a tighter relaxation for our problem is hence to demand the existence of a state of the form denoted in Figure 7.1.

There Alice and Bob's quantum systems are represented by diamonds; we will call such systems *heads*. Circles joined by a line will be called *legs*; they represent Alice and Bob's rank-one projectors. The number of circles in a leg will be the *length* of the leg. Mathematically, a diagram like the above represents a particular relaxation of the separability condition. Namely, drawing a leg of length N for a particular measurement will indicate that we will be approximating the associated rank-one projector by a subsystem of an N-symmetric ensemble. In principle, one can demand the positivity of the whole state under the partial transposition of any number of circles (not necessarily belonging to the same leg). Such a condition will be denoted by encompassing with a dashed line the relevant circles.

This time, the action of the operators M_a^x (N_b^y) can only be non-trivial in one of the circles of leg x (y) and Alice's (Bob's) head. By [DPS05], as we increase the length of the legs on the diagram, we converge to the solution of problem (7.-2).

Informally speaking, the idea behind the method so far is to force the tensor product structure between state and measurements through the N extension of the measurements (legs) and the positive partial transpose conditions. The longer the length, the closer to a tensor product.

Higher dimensions and higher number of outcomes

In order to optimize dichotomic Bell inequalities over higher dimensional Hilbert spaces, again we can assume that measurements are projective. This time, though, there may exist non-trivial projectors with rank greater than 1. To model a rank-2 projector, we must then introduce two legs, one for each rank-1 projector, and then enforce orthogonality relations between them. Denoting by C, D the circles of two different legs, the orthogonality condition is translated as

$$\mathrm{tr}_{CD}\{W \cdot V(C,D)\} = 0. \tag{7.-3}$$

Similar considerations apply to optimizations involving d-valued projective measurements.

As for the simulation of generalized measurements with more than two outcomes, note that any POVM can be viewed as a projective measurement in a larger Hilbert space. Namely, for any set of POVM elements $\{M_a\}_{a=0}^{A-1} \subset B(\mathbb{C}^d)$, there exists a complete set of projectors $\{\Pi_a\}_{a=0}^{A-1} \subset B(\mathbb{C}^{d'} \oplus \mathbb{C}^d)$, such that $M_a = (0_{d'} \oplus \mathbb{I}_d)\Pi_a(0_{d'} \oplus \mathbb{I}_d)$ for $a = 0, ..., d-1$. Hence, in order to play with d-outcome generalized measurements it suffices to consider projective measurements in a larger Hilbert space, project them into the original space and collapse them with Alice's or Bob's head, depending on the case. The amount of resources needed increases very quickly with the number of measurement outcomes, though.

7.2.2. Hybrid infinite-finite dimensional optimization

As we saw in the last section, with the previous approach, when we enforce the PPT condition (and thus are bound to use SDP), even Bell optimizations in simple scenarios like the 4422 are intractable with a normal desktop. In this section we will improve the previous algorithm to deal with scenarios where just one of the parties has many measurement outcomes. As we will see, the new

algorithm can be extended straightforwardly to deal with multipartite situations where the local dimensions of a subset of the parties are constrained, while the rest have access to infinite dimensional degrees of freedom.

The key to this improvement is a process that we will denominate 'body expansion'.

Body expansion

For simplicity, picture a tripartite scenario where one of the parties, Alice, has total control over her d-dimensional quantum system living in the Hilbert space A, but we completely ignore the operations being carried out by the other two observers, call them Bob and Charlie, in their Hilbert spaces B and C. That is, we are contemplating a nonlocality scenario where the measured correlations are of the form

$$P(a,b,c|x,y,z) = \mathrm{tr}(\rho_{ABC}\Pi_a^x \otimes E_b^y \otimes F_c^z), \tag{7.-3}$$

where $\{\Pi_a^x\} \subset B(\mathbb{C}^d)$, acting in A, are known measurement operators, and $\{E_b^y, F_c^z\}$, acting in B and C respectively, represent unknown projector operators acting over arbitrary Hilbert spaces.

Based on the local mapping approach introduced by Moroder *et al.* [MBL$^+$13], Pusey [Pus13] recently proposed to characterize this class of systems by expanding the unknown degrees of freedom in a moment matrix *à la NPA* [NPA07] while keeping the trusted system the same. This notion can also be found in prior work by Helton & McCullough [HM04], but, for didactical purposes, we will follow Moroder *et al.*/Pusey's presentation.

Given the multipartite state ρ, the idea is to implement the map

$$\rho \to \mathrm{tr}_B(\mathbb{I}_A \otimes \Lambda_{BC})\rho(\mathbb{I}_A \otimes \Lambda_{BC})^\dagger, \tag{7.-3}$$

with

$$\Lambda_{BC} = \sum_{|s|\leq n} s \otimes |s\rangle\,. \tag{7.-3}$$

Here the sum is over all sequences s of unknown projectors $\{E_b^y, F_c^z\}$ of length $|s|$ smaller than or equal to n (including the identity), and $\{|s\rangle\}$ is an orthonormal basis where each vector is labeled by a sequence of $\{E_b^y, F_c^z\}$. Defining $c_s^{k,j} \equiv \mathrm{tr}\{(|j\rangle\langle k| \otimes s)\rho\}$, it can be seen that the result of such a map is a positive semidefinite operator of the form

$$\Gamma^{(n)} \equiv \sum_{k,j} |k\rangle \langle j|_A \otimes \sum_{|s|,|t|\leq n} c_{t^\dagger s}^{k,j} |s\rangle \langle t| , \tag{7.-3}$$

with

$$\sum_k c_{\mathbb{I}}^{k,k} = 1. \tag{7.-3}$$

From now on, the matrix $\Gamma^{(n)}$ will be called a *generalized moment matrix*. It is worth noting that here we are identifying sequences of operators modulo commutation relations, i.e., $E_b^y F_c^z$ and $F_c^z E_b^y$ are regarded as the same sequence. Also, 'null sequences' like $E_b^y E_{b'}^y$, with $b \neq b'$, are not considered, or, equivalently, their corresponding coefficients $c_s^{k,j}$ are set to zero.

If we use (7.2.2) rather than (7.2.2) to represent the quantum systems involved in the experiment, we will say that the body of parties B, C has been *expanded*. The original probability distribution can be retrieved by

$$P(a,b,c|x,y,z) = \mathrm{tr}\{\Gamma^{(n)}(\Pi_a^x \otimes |t\rangle \langle s|)\}, \tag{7.-3}$$

where t, s are any two sequences such that $|s|, |t| \leq n$ and $t^\dagger s = E_b^y F_c^z$. It is straightforward to extend the notion of body expansion to more than two parties.

Actually, due to the linear dependence $E_{\tilde{b}}^y = \mathbb{I} - \sum_{b \neq \tilde{b}} E_b^y$, it is enough to consider sequences of projector operators corresponding to the first $d-1$ outcomes in eq. (7.2.2). This allows saving computer memory and leads to the same numerical results, so from now on we will be assuming that generalized moment matrices are only defined on such sequences.

In general, demanding the existence of a positive semidefinite operator $\Gamma^{(n)}$ of the form (7.2.2) constitutes a relaxation of the original problem of characterizing the convex hull of all distributions of the form (7.2.2). Hence, in order to achieve convergence, we must consider a hierarchy of semidefinite programs $\Gamma^{(1)} \geq 0, \Gamma^{(2)} \geq 0, ...$, see [Pus13, HM04]. However, in the case where just one of the parties was expanded, it is enough to impose $\Gamma^{(1)} \geq 0$ (see Appendix A.1).

Expanded bodies in dimension-bound Bell scenarios

Consider a tripartite Bell scenario where the local dimension of one of the parties is bounded: the situation is similar to that in the previous section, i.e., eq. (7.2.2) holds. This time, however, we ignore the mathematical expression of Alice's measurement operators $\{\Pi_a^x\}$. Our solution is, of course, to combine the

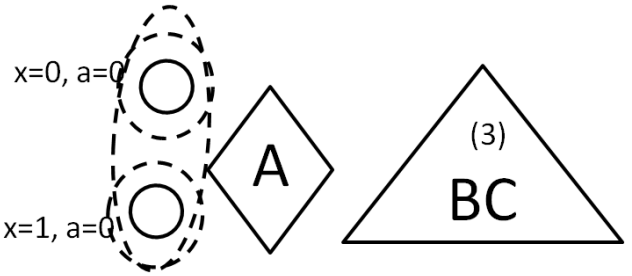

Figure 7.2.: A possible relaxation to bound the convex hull of eq. (7.2.2) when Alice's measurements are not trusted. Here just Bob and Charlie have expanded bodies, represented by a triangle. The number (3) indicates the order of the moment relaxation.

body expansion technique to deal with dimension unconstrained systems with the Method developed in Section 7.2.1 to deal with dimension constraints on the quantum system of some party. Figure 7.2 shows a diagrammatic representation of a possible relaxation for this problem.

This diagram has to be understood as follows: the triangle represents the Hilbert space defined by $\text{span}(|s\rangle : |s| \leq 3)$ in eq. (7.2.2); the rest of the figures represent the first Hilbert space in the same expression. This first Hilbert space can be expressed as a tensor product of three Hilbert spaces (Alice's head and her two legs). Note as well that the positivity of the partial transpose of the matrix (7.2.2) with respect to three different subsets of Alice's legs is being enforced. Further -better- relaxations are attained by increasing the order n of the expansion, and the length of Alice's legs.

In such a general case, probabilities are extracted from the main matrix via the formula:

$$P(a,b,c|x,y,z) = \text{tr}\{(M_a^x \otimes |t\rangle\langle s|)\Gamma^{(n)}\}, \tag{7.-3}$$

where M_a^x are, recall, the operators containing the SWAP operator as defined in (7.-1), and s,t are any two sequences such that $t^\dagger s = E_b^y F_c^z$.

It can be shown (see Appendix A.2) that the above method converges to the convex hull of the set of all distributions of the form

$$P(a,b,c|x,y,z) = \text{tr}\{(\Pi_a^x \otimes E_b^y F_c^z)\rho\}, \tag{7.-3}$$

with Π_a^x (E_b^y, F_c^z) acting over a d-dimensional (finite or infinite dimensional) Hilbert space, and $[E_b^y, F_c^z] = 0$.

When E_b^y, F_c^z act over a finite dimensional Hilbert space, the above expression can be proven equivalent to eq. (7.2.2) [SW08], i.e., one can identify commutativity and tensor products. In the infinite dimensional case, though, this is no longer the case, and the existence of a tensor representation for (7.2.2) relies on the validity of Kirchberg's conjecture, a major open problem in mathematics [Fri12, JNP+11]. It shall be noted that this technical limitation, already present in the NPA hierarchy [NPA08], only concerns the *convergence* of the SDP schemes presented in this section. That is, independently of whether Kirchberg's conjecture is true or not, the algorithms proposed above constitute a rigorous relaxation of the original tripartite characterization problem.

Remark 1. Suppose that we wish to characterize the set of bipartite distributions

$$P(a,b|x,y) = \mathrm{tr}(\Pi_a^x \otimes E_b^y \rho_{AB}), \tag{7.-3}$$

with $\{\Pi_a^x, E_b^y\}$ acting over $B(\mathbb{C}^d)$ but otherwise unknown. From the Schmidt decomposition, this scenario can be seen equivalent to just limiting Alice's operators to act over $B(\mathbb{C}^d)$, while allowing Bob's operators to access Hilbert spaces of arbitrarily high (or even infinite) dimension. Hence we can expand Bob's body to the first order while assigning head and legs to Alice. From Appendix A.1, it follows that expanding Bob's head to higher orders will not improve the approximation; convergence to (1) is thus achieved simply by increasing the length of Alice's legs. Note that the size of the corresponding generalized moment matrix is still exponential in Alice's number of measurements, but linear in Bob's. With this trick, the 4422 scenario, as well as others of the form $4m22$, can therefore be optimized on a normal computer.

7.2.3. Application examples

As an application of the techniques developed in the preceding sections, we provide several examples.

First we discuss dimension witnesses for the 2-party scenario. These witnesses are actually Bell-type inequalities whose violation gives a lower bound on the dimension of the Hilbert space. We first apply the entanglement based method of Section 7.2.1 for witnessing dimension in two-party systems using three-setting dichotomic Bell inequalities.

Then we move to more demanding Bell inequalities with Alice having four dichotomic settings and Bob having up to twelve dichotomic settings by using the method of Section 7.2.2. Next we discuss the multipartite case by fixing the local Hilbert space of one of the parties to be two dimensional, but we

do not impose any bound on the dimension of the rest of the parties. This hybrid scenario will allow us to certify true three-dimensional entanglement in a device-independent manner. For this sake, we make use of a three-party Bell inequality having three dichotomic settings per party, which turns out to be a minimal construction. Finally, it is demonstrated that our technique is also suitable to certify entangled measurements in finite dimensional Hilbert spaces in a rigorous way.

In the following computations we used the MATLAB package YALMIP [Lof04] and the SDP solvers SeDuMi [Stu99], CSDP [Bor99], SDPLR [BM03] and SDPNAL [ZST10].

Two-party dimension witnesses

A family of three setting dimension witnesses

Let us consider the tilted version of the so called I_{3322} inequality. The $I_{3322} \leq 0$ can be defined as

$$I_{3322} = -P(A_1) - \sum_{y=2}^{3} P(B_y) + \sum_{y=1}^{3} P(A_1, B_y) + \sum_{x=2}^{3} P(A_x, B_x) - \sum_{1 \leq y < x \leq 3} P(A_x, B_y)$$

where the notation is such that $P(A_x, B_y) = P(0, 0|x, y)$. The tilted version is a modification on the original inequality in order to deal with the situation in which Alice has a perfect detector but Bob's detector only works with probability η. A violation on the tilted inequality means that the original inequality tolerates a detection efficiency η on Bob's side. The tilted inequality is $I_{3322}(\eta) = I_{3322} - \frac{1-\eta}{\eta} P(A_1)$. This one-parameter family of inequalities $I_{3322}(\eta) \leq 0$ is parametrized by $1/3 \leq \eta \leq 1$. We refer the reader to the references [BGSS07, VPB10] for details on this family of inequalities. Strong numerical evidence shows [VPB10] that this inequality cannot be violated by conducting measurements on qubits if $\eta \leq 0.428$. Using the technique of Section 7.2.1, we show that the limit is indeed $\eta \simeq 0.428$ subject to numerical precision of the SDP solver SeDuMi [Stu99]. In Table I we present the two-qubit maximum results for various η values. In particular, the lower bound value arises from a see-saw iteration procedure [PV10], where all respective measurements turn out to be on the X-Z plane. The upper bound value, on the other hand, is due to the SDP technique of Section 7.2.1. Note that by $\eta \simeq 0.429$ the SDP upper bound value becomes comparable with the precision of our SDP solver ($\sim 10^{-9}$). As it can be observed, the lower-bound and upper-bound values are in good agreement for $\eta \geq 0.45$. The complexity of the SDP problem can be characterized by the number of constraints involved and

η	Lower bound	Upper bound
1	0.25000	0.25000
0.8	0.14331	0.14331
0.6	0.03910	0.03910
0.5	0.00608	0.00608
0.45	2.8014×10^{-4}	2.8015×10^{-4}
0.44	4.8213×10^{-5}	5.8207×10^{-5}
0.43	1.0764×10^{-7}	8.7542×10^{-7}
0.429	2.9466×10^{-9}	5.9880×10^{-8}
0.428	$\sim 10^{-17}$	3.7484×10^{-9}

Table 7.1.: Lower and upper bounds on the violation of the $I_{3322}(\eta)$ inequality in the two-qubit Hilbert space. The local bound is equal to 0 for any η displayed.

the dimension of the underlying semidefinite matrix. In our particular case, the respective numbers are 2080 and 1027, and solving the SDP problem took about 1 minute on a desktop PC.

Four setting dimension witnesses

The technique presented in Section 7.2.2 is computationally cheaper than the one of Section 7.2.1 used previously for three setting inequalities. So let us utilize this more powerful technique to construct dimension witnesses with four measurement settings per party. Firstly consider a four setting tight Bell inequality, which is the $N = 4$ member of the I_{NN22} family [CG04]. Here a qubit lower bound is given by 0.25, when Bob measures a rank-0 projector in one of his settings (and the rest of the measurements are rank-1 projectors). Note that a rank-0 projector accounts for a never-occurring outcome of a measurement. In the following, we will call such measurements *degenerate*. This value of 0.25 could not be overcome using the see-saw variational technique. The qubit upper bound due to our SDP algorithm is given by 0.26548, whereas the maximum overall quantum value certified by the NPA hierarchy is 0.28786, which is attainable with real-valued qutrit systems [PV09b]. Hence, a Bell violation bigger than 0.26548 serves as a dimension witness, signaling the presence of qutrit systems. In the present case, the number of constraints involved in the SDP problem is 3241 and the dimension of the semidefinite matrix is 883. Our desktop PC required about 15 minutes to solve the problem.

Another inequality, which is not tight but despite its simplicity gives a di-

mension witness with relatively good noise tolerance, is the following one:

$$\begin{aligned} I_{4,4} =& E_1^A + E_{1,1} + E_{1,2} + E_{2,1} - E_{2,2} \\ &+ E_{3,3} + E_{3,4} + E_{4,3} - E_{4,4} \leq 5, \end{aligned}$$

where the correlator $E_{x,y}$ between measurement x by Alice and measurement y by Bob is defined as $E_{x,y} = P(a = b|x, y) - P(a \neq b|x, y)$, and E_x^A denotes the single-party marginal of Alice's x-th measurement setting. Notice that the inequality is composed by a CHSH inequality (for settings 3 and 4) and a tilted CHSH inequality (for settings 1 and 2). An upper bound is given by adding up the maximum quantum value of these two Bell expressions [AMP12b], $Q = 2\sqrt{2} + \sqrt{10} \simeq 5.9907$. This bound can in fact be saturated by a 2-ququart system, by tensoring a 2-qubit singlet state with a 2-qubit partially entangled state. However, if we fix dimension two for the Hilbert space of both parties, we expect not to attain the overall quantum maximum. Indeed, numerical evidence shows that for qubits the limit is 5.8310, whereas the upper bound using the expanded bodies technique of Section 7.2.2 is given by 5.8515. Hence, a value bigger than 5.8515 certifies three-dimensional systems. We tried to increase the order of the expansion in order to get even better upper bounds in both above cases, but unfortunately the SDP problem was not feasible using the solvers SeDuMi [Stu99] or CSDP [Bor99] on a normal desktop computer.

Correlation type dimension witnesses

We investigate the qubit bound of correlation type Bell inequalities, where Alice has four and Bob has up to twelve dichotomic measurement settings. We consider the following linear functions of correlators $E_{x,y}$,

$$I_{m_A,m_B} = \sum_{x=1}^{m_A} \sum_{y=1}^{m_B} M_{x,y} E_{x,y} \leq L, \tag{7.-4}$$

where m_A and m_B are the number of settings on Alice and Bob's side, respectively. Hence, I_{m_A,m_B} defines an (m_A,m_B) setting correlation type Bell inequality, where L denotes the local bound. Let's take three such Bell inequalities, defined by the coefficient matrices M as follows,

$$M_{4,7} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & -1 & -1 \end{pmatrix}, \tag{7.-4}$$

and

$$M_{4,8} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}, \tag{7.-4}$$

and

$$M_{4,12} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 1 & -1 \end{pmatrix}. \tag{7.-4}$$

The local bound of the corresponding inequalities $I_{4,7}$, $I_{4,8}$ and $I_{4,12}$ are given by 8,12, and 12, respectively. Note that all above Bell inequalities are members of a larger family [VP08]. In particular, $I_{4,8}$ is a straightforward generalization of Gisin's elegant inequality [Gis07].

Applying the method of Section 7.2.2 for the case of two parties, we get the two-qubit upper bounds summarized in Table II. As a comparison, the qubit lower bound and ququart maximum values are also given. According to the table, each three inequalities serve as dimension witnesses. Note, however, that there are small gaps between the upper and lower bounds obtained. Hence we programmed a higher relaxation, depicted in Figure 7.3, of the method of Section 7.2.2 to bound the value of $I_{4,12}$, which allowed us to close the gap. To implement this second order relaxation, we used a memory-enhanced desktop and the SDP solver SDPNAL [ZST10]. This case was the most demanding among all the studied examples from a computational point of view. Here the number of constraints was 1385281 and the dimension of the underlying semidefinite matrix was 13312 and took about 13 hours for our computer to solve the problem.

As a side note, let us mention that in the present case of correlation type Bell inequalities, it was enough to consider rank-1 projective measurements (i.e., no need to take into account degenerate measurements), since this type of inequalities is known to be maximized in the two-qubit space by using rank-1 projective measurements [AGT06].

Genuine tripartite higher dimensional entanglement

Let us consider a three party three setting Bell inequality, which is invariant under any permutations of the three parties and it has the peculiarity that it consists of only 2-party correlation terms, which is usually an advantage in

	Qubit lower bound	Qubit upper bound	Ququart
$I_{4,7}$	10.4995	10.5102	10.5830
$I_{4,8}$	15.4548	15.7753	16
$I_{4,12}$	16.7262	16.7645 (16.7262)	16.9706

Table 7.2.: Qubit lower/upper bounds on the violation of the I_{m_A,m_B} inequalities defined by Eqs. (7.2.3,7.2.3,7.2.3) computed using see-saw iteration/derived from our construction. The number between brackets in the second column corresponds to the SDP relaxation of Figure 7.3. The ququart value defines the overall quantum maximum given by the zeroth level of the NPA hierarchy.

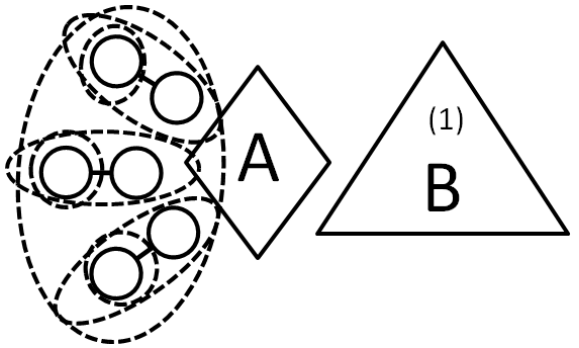

Figure 7.3.: Pictorial representation of the second relaxation used to compute the maximal violation of $I_{4,12}$ in qubit systems. The three measurement's on Alice side live on the symmetric space of 2 particles. We used the freedom of the problem to fix the fourth measurement of Alice. Positivity under the partial transposition of different combination of Alice's legs has been imposed. The number (1) on Bob's head indicates that it has been expanded to first order in its moment relaxation since as proved in Appendix A.1 this is enough for convergence

experimental situations. The inequality is as follows:

$$\begin{aligned} I_{3,3,3} =& \text{sym}\{-P(A_1) - 2P(A_3) + P(A_1, B_1) \\ & - 2P(A_2, B_2) - 2P(A_3, B_3) - P(A_1, B_2) \\ & + P(A_1, B_3) + 2P(A_2, B_3)\} \leq 0, \end{aligned}$$

where $\text{sym}\{X\}$ denotes that all terms in the expression X have to be symmetrized with respect to all permutations of the parties, and we used the simplified notation $P(A_x, B_y, C_z) = p(0, 0, 0|x, y, z)$. On one side, we computed lower bound values arising from the heuristic see-saw search for different dimensionalities of the parties, $2 \times 2 \times 2$, $2 \times 3 \times 3$ and $3 \times 3 \times 3$. Note that the cases $2 \times d \times d$, $d \times 2 \times d$, $d \times d \times 2$ refer to the same situation, because the inequality (7.2.3) is fully symmetric. Therefore, it is enough to perform optimization in one of the cases, say, $2 \times d \times d$, where dimension $d \geq 2$. On the other side, we give the upper bound value for the case of $2 \times \infty \times \infty$ (that is, when Alice acts on qubits, and the other parties have no restriction on the dimension). Due to the symmetry of the inequality, again, the same upper bound applies to the $\infty \times 2 \times \infty$ and $\infty \times \infty \times 2$ situations as well.

In the present case, we have to take into account degenerate measurements (either rank-0 or rank-2 projective measurements) on Alice's side, in which case the inequality (7.2.3) reduces to a two setting inequality on Alice's side, hence Alice's qubit state space suffices to obtain maximum quantum violation [Mas05]. That is, when we compute an upper bound on $2 \times \infty \times \infty$ in the degenerate case, we can use the dimension unrestricted case of the NPA method [NPA07]. Table III summarizes the results obtained. By eye inspection, both upper bounds are saturated, hence they are tight (up to numerical precision). Hence, any Bell violation of $I_{3,3,3}$ bigger than 0.1786897 witnesses in a device-independent way that the underlying state ρ_{ABC}, not only has Schmidt number vector $(3, 3, 3)$ [HdV13], but also that any pure state decomposition of ρ_{ABC} contains at least one state $\sigma_{ABC} = |\psi\rangle\langle\psi|$ such that $\text{rank}(\sigma_A), \text{rank}(\sigma_B), \text{rank}(\sigma_C) \geq 3$. To illustrate the power of this Bell inequality, let us pick the following state

$$|\Psi\rangle = \cos\alpha\,|\psi\rangle + \sin\alpha\,|222\rangle\,, \tag{7.-6}$$

with $\alpha = 0.2519038$, where $|\psi\rangle = (|012\rangle + |021\rangle + |102\rangle + |120\rangle + |201\rangle + |210\rangle)/\sqrt{6}$ is the fully (bosonic) symmetric 3-qutrit state. By optimizing over the measurement angles in the X-Z plane, we get the quantum value $Q = 0.1841287$. Since this value is clearly bigger than the threshold 0.1786897, we can argue device-independently that the above state (7.2.3) is genuinely three-dimensional.

	LB (222)	LB (233)	UB (2∞∞)	LB (333)
No-deg	0.0443484	0.1783946	0.1783946	0.1962852
Deg	0.1783946	0.1786897	0.1786897	

Table 7.3.: Qubit lower/upper bounds for different local dimensions on the violation of the $I_{3,3,3}$ inequality computed using see-saw search/SDP computation. Qutrit value (333) is the overall quantum maximum as certified by the NPA hierarchy. The upper bound value for the non-degenerate case (denoted by No-deg) was computed using the technique of Section 7.2.2, whereas the upper bound value for the degenerate case (denoted by Deg) was obtained by the NPA hierarchy. Abbrevation LB/UB refers to lower/upper bound.

Entangled measurements in two-qubit Hilbert spaces

Let us consider the following scenario, pictured in Fig 7.4: two separated parties, Alice and Bob, have each a preparation device which prepares unknown qubit states out of 3 possible respective states ρ_x and σ_y. These states are sent to Charlie's two distinct ports C_A and C_B, who in turn interacts with the received states and announces a bit c. The experiment is described by a set of conditional probabilities $P(c|x,y) = \mathrm{tr}\,(\rho_x \otimes \sigma_y M_c)$, where $M_c, c = 0, 1$ denote Charlie's POVM elements.

Depending on the form of M_c one can distinguish between different scenarios. In case of *unentangled measurements*, each of the POVM elements M_c is a separable operator. Moreover, it is known that a subclass of this class corresponds to *LOCC measurements*, in which case M_c is associated with a sequence of measurements on C_A and C_B ports, with each measurement depending on the outcomes of earlier measurements. On the other hand, in case of *general measurements*, the measurement operators in quantum mechanics are only limited by positivity and normalization, and they can be well entangled. For instance, Bell state measurements belong to this class.

We consider the following witness, introduced in [VN11]:

$$W = -P_{11} - P_{12} + P_{13} + P_{21} + P_{23} + P_{31} - P_{32} - P_{33}, \qquad (7.\text{-}6)$$

where we identify $P_{xy} = P(0|x,y)$. Using a see-saw type iteration, we obtained the bound $w_{gen} = 2.5$ for general measurements and $w_{unent} = (2 + 3\sqrt{6})/4 \simeq 2.3371$ [VN11]. Note, however, that due to the heuristic nature of the see-saw type search, these bounds are not rigorous, they constitute only a lower bound to the problem. On the other hand, adapting the technique of Section 7.2.1

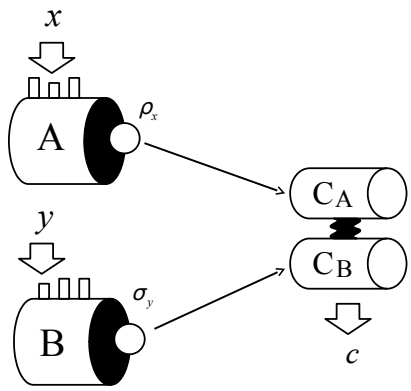

Figure 7.4.: Picturing the scenario of entangled measurements in bounded Hilbert spaces.

to the present case, we get an upper bound of 2.506 for w_{gen} (in this case, the solver SDPLR was used [BM03]). In the unentangled case, we identify separable measurements with rank-2 projectors, which may not be justified in general. However, modulo this condition, we get the upper bound of 2.3371 for w_{unent}. In the latter case, rank-2 projective measurements are composed by the sum of two orthogonal rank-1 projectors. Then, we have to define a leg for each rank-1 projector, and impose that the legs are orthogonal, as described in Sec 7.2.1.

Note that the result $w_{gen} < 2.506$ allows us to turn around the problem. Namely, suppose that there is no dimensionality constraint on Alice and Bob's emitted states ρ_x and σ_y. Then the inequality $w_{gen} \leq 2.506$ may work as a dimension witness: its violation guarantees that at least qutrits had to be prepared by Alice or Bob (or by both parties).

7.3. Discussion

In this fourth Chapter we have studied the central problem in semi-device-independent quantum information: bounding the strength of quantum nonlocality under local dimension constraints. By relating finite dimensional quantum correlations to the separability problem, we have managed to exploit existing entanglement detection criteria to devise hierarchies of SDP relaxations

for the characterization of quantum nonlocality in multipartite scenarios with a promise on the local dimensions of the parties involved. The first relaxations of our method were applied successfully to upper bound the maximal violation of several bipartite Bell inequalities in qubits. The relatively small memory resources required to implement our method allowed us to investigate with a normal desktop bipartite Bell scenarios with 4 settings on one side and 12 on the other. Although it was not always possible to close the gap between our upper bounds and the corresponding lower bounds obtained via variational methods, our SDP relaxations output results below the quantum maximum in all considered cases. Moving on to tripartite scenarios, we applied the method to identify a tripartite Bell inequality that cannot be violated maximally if any one of the parties holds a qubit. This inequality can hence be used to certify device-independently that a tripartite quantum state has genuine three dimensional entanglement [HdV13], showing how bring to the black box realm the entanglement characterization of [HdV13]. Finally, we applied the hierarchy to certify entangling measurements in two-qubit Hilbert spaces, as in [VN11].

The reader may have noted that, no matter the dimension of the local Hilbert space, all our examples involved dichotomic measurement operators. The reason is that extremal dichotomic measurements are known to be projective, and so they admit a simple representation in terms of legs. The characterization of many-outcome POVMs is, however, not so straightforward. In principle any POVM can be expressed as a projective measurement in a higher dimensional Hilbert space, and so our method can be adapted, via dimension enhancement, to Bell scenarios involving more than two outcomes. However, the known bounds on the minimal dimension required for arbitrary POVMs are high enough as to make our method impractical in a normal computer, see [CBZG07]. It is an open question whether extreme many-outcome POVMs require considerably less dimension resources, like in the qubit case [DPP05] or, more generally, whether the description of extremal POVMs can be simplified to the point of making our new method feasible for such Bell scenarios.

8. Implications of Church-Turing thesis for quantum physics

The Church-Turing thesis is one of the pillars of computer science; it postulates that every classical system has equivalent computability power to the so-called Turing machine.

In this Chapter we analyse some consequences of this thesis in the context of quantum physics and show that computer science laws have implications for some of the most fundamental results of the theory. Firstly, we will show how different preparations of the same mixed state, undistinguishable according to the quantum postulates, when prepared by a computer turn out to be distinguishable. Secondly, we also show a new loophole for Bell-like experiments: if some of the parties in a Bell-like experiment use a computer to decide which measurements to make, then the computational resources of an eavesdropper have to be limited in order to have a proper observation of non-locality. The results presented on this Chapter are based on the article [BdlTS$^+$14].

8.1. Introduction

Most of the effort until now studying the relationship between quantum mechanics and the Church-Turing thesis has concerned on how the former can affect the latter (see, for instance, [AD12]). In this Chapter, however, we take an opposed approach showing that the Church-Turing thesis must be taken into account in order to avoid apparent paradoxes.

Let us now put our two results in that direction into context. First we show that computers impose a limitation when it comes to producing a mixed state as a classical mixture of pure quantum states. It turns out that with the sole knowledge that the classical mixture is performed by a computer, situations that seem not to be distinguishable turn out to be so. This has direct implications since mixed states are prepared this way in many experiments [AB09, LKPR10]. Secondly, when it comes to Bell-like experiments to test non-locality, another distinctive feature of quantum mechanics, we show that if the measurement independence between the two parties [CK06, KHS$^+$12, Hal10, BG10] is achieved via private computable pseudo random number generators, an eavesdropper,

even one without any previous knowledge, can start guessing their inputs from the information on their previous inputs, thus leading to a new computability loophole for Bell tests.

Formally, our results apply only to computers. This has already considerable practical consequences, since almost every experimental setup is controlled by classical computing devices. However, one can argue that, due to the widely accepted physical interpretation of the Church-Turing thesis:

> *the behaviour of any discrete physical system evolving according to the laws of classical mechanics is computable by a Turing machine,*

our results, in fact, apply to every classical system and hence, the limitations that we show are fundamental.

8.2. Results

8.2.1. Proper mixed state preparation and the Church-Turing thesis

We start by considering one of the basic parts of quantum theory: the concept of a mixed state [NC00]. We will see that although well understood by physicists, their nature and origin can lead to apperent paradoxes when confronted against common computer science tenets.

Let us present now the following preparation of a mixed state: A classical computer in an unknown configuration and with unbounded memory is running an unknown and presumably very convoluted algorithm to prepare a mixed state. We have the promise that the computer, understood as a black box, is mixing evenly either the single qubit eigenstates of σ_z $\{|0\rangle, |1\rangle\}$ or the states $\{|+\rangle, |-\rangle\}$, where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ as seen in Fig. 8.1. Is there any operational procedure to decide which of the two ensembles are being mixed for an experimenter (Bob) who cannot open the black box? Even though one would be tempted to assign to both preparations the identity state $\rho = \frac{\mathbb{I}}{2}$, our results show that the fact that the mixing procedure was performed in a computable way leaves a trace which allows us to distinguish both mixtures in finite time and with arbitrarily high success probability. It is worth mentioning that having a computer mixing the state doesn't imply that the sequence in which it mixes the state is periodic. In fact, there exist normal sequences (i.e. those which satisfy the law of large numbers in a generalized way), or other even 'more random' sequences which are computable in polynomial time [FN13].

In order to solve this problem, Bob measures every qubit that comes out of the black box on an odd position in the basis of eigenstates of σ_z, yielding a binary sequence of measurement results Z. He also measures every qubit on an

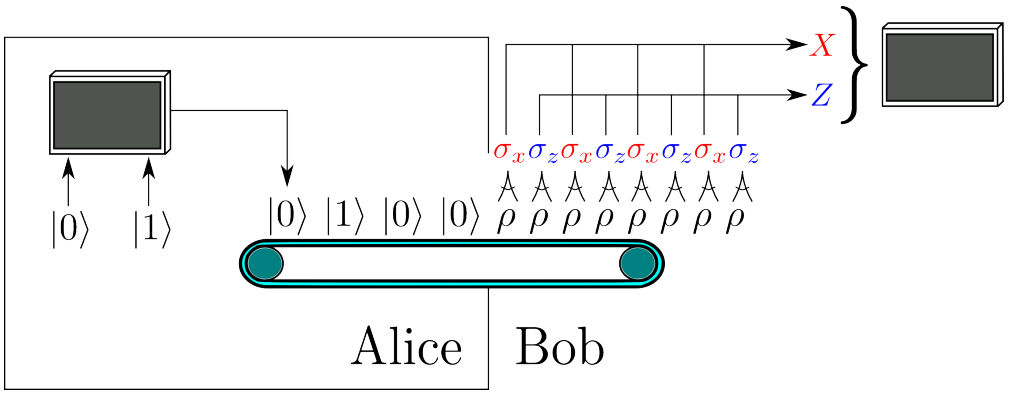

Figure 8.1.: Scheme for the preparation of mixed states via classical mixing. Alice uses a computer to choose between $|0\rangle$ and $|1\rangle$ (or $|+\rangle$ and $|-\rangle$). From Bob's perspective, since he doesn't know the computer program Alice is using, it's a mixed state ρ. To distinguish both possible preparations he measures alterantively σ_x and σ_z and sends the resulting sequences to a computer that will be able to tell which of the sequences is computable (corresponding to the basis in which Alice prepared the state) and which is a fair coin tossing (corresponding to the other basis).

even position in the basis of eigenstates of σ_x obtaining a binary sequence X. This way Bob obtains two binary sequences, as can be seen in Fig. 8.1. The one corresponding to the choice of measurement that matches the preparation basis is computable, and the other one corresponds to a fair coin tossing. Therefore we need an algorithm that given two sequences, one computable and one arising from a fair coin tossing, is able to tell us which is which. We will show now one such algorithm, that can perform this task in finite time and with an arbitrarily high probability of success.

To distinguish which of the two sequences is computable we dovetail between program number (the programs are computably enumerable) and maximum time steps that we allow each program to run (that is, we run program 1 for 1 timestep, then programs 1 and 2 for 2 timesteps and so on), as is a common technique in computability theory. For each program p of length $|p|$ we will compare the first $k|p|$ output bits with the corresponding prefixes of both sequences, where k is an integer constant depending on the probability of success we are looking for. Whenever we find a match for the first $k|p|$ bits, we halt. Fig. 8.2 depicts the dovetailing algorithm. It is straightforward to see that this algorithm always gives an answer. That the probability of making

a mistake is less than $O(2^{-k})$ will be shown in appendix E.1. Therefore we can guess in finite time and with an arbitrarily high probability of success (by setting k we adjust the probability of success).

The complete algorithm to distinguish a fair coin from a computable sequence is Algorithm 1 below, where $X \upharpoonright k|p|$ denotes the fist $k|p|$ bits of the sequence X.

Algorithm 1 The distinguishing protocol. $U_t(p)$ is a universal Turing machine that runs program p for t timesteps. The two 'for' loops correspond to the dovetailing.

Require: $k \in \mathbb{N}$ and $X, Z \in 2^\omega$, two bit sequences with the promise that one of them is computable.

Ensure: 'X' or 'Z' as the candidate for being computable; wrong answer with probability bounded by $O(2^{-k})$.

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

Note that, at a given iteration of the algorithm, it may perfectly be the case that program p has not been able to produce in t time steps the $k|p|$ symbols needed to check the halting condition. If this is the case, the algorithm simply keeps running and moves to the next program. However, the algorithm will surely halt as it will run the actual program used in the blackbox at some finite time. For a detailed explanation on how Algorithm 1 works and its probability of success, see Appendix E.1.

It is an interesting open question to study the effect of noise in the previous algorithm. As a first step, we have considered a rather simple noise model in the state preparations and measurements described by a flip probability in the observed symbols r. That is, we consider the situation in which those results obtained when measuring the quantum states in the actual basis used by the box are correct with probability $1 - r$ (this simple noise has no effect on the results of measurements performed in the wrong basis). As shown in the Appendix E.1, there is another slightly more complex algorithm that still halts

with arbitrarily small error probability whenever $r \lesssim 0.21$. The key idea is to change the halting condition: instead of looking for a program that generates the prefix of either of the sequences, we look for a program that generates a prefix with at most a fraction r of bits flipped from either of the sequences.

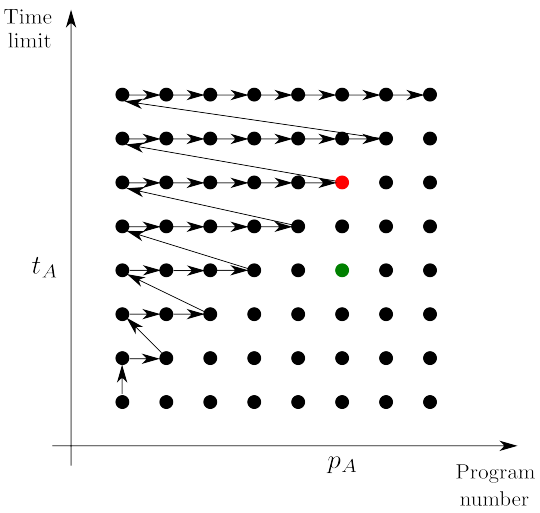

Figure 8.2.: To decide which of the two sequences is computable, we dovetail between program number and timesteps that each program is allowed to run. In green we show the actual program p_A that was used to prepare the state, and the number of timesteps t_A it takes to generate a $k|p_A|$ long prefix. The latest halting condition for our algorithm is shown in red, although it might halt before that with either a wrong recognition or a correct one.

The previous algorithm is of course very demanding, but proves that classical mixtures of pure states that seem to define the same mixed state are in principle distinguishable when prepared by classical computing devices. This leaves quantum mixtures (either by using a part of a larger entangled system or a quantum random number generator) as the only adequate way of creating mixed states. It is important to point out that, although the algorithm to distinguish will tipically take an extremely long time, it is just a tool to show that the two situations are *distinguishable*, unlike what is expected if one assumes that some computable pseudorandomness is enough to prepare such states. Our results imply that non-computable resources are needed to produce proper mixed states. Alternatively, assuming the validity of both quantum theory and the Church-Turing thesis, the only systems correctly described by a mixed state density matrix are those forming part of a bigger quantum system

in a pure state.

8.2.2. Bell test computability loophole

As we have seen in detail throughout this thesis, non-locality is another of the most intrinsic features of quantum mechanics [EPR35, Bel64, Bel66]. Recall that the standard Bell scenario is described by two distant observers who can perform m possible measurements of d possible outputs on some given devices. The measurements are arranged so that they define space-like separated events. For the sake of simplicity we focus our attention in the standard bipartite 2-input 2-output scenario [Bel66, CHSH69], although our considerations apply to any Bell scenario. It is convenient for what follows to rephrase the standard Bell scenario in cryptographic terms, as in [BCH$^+$02, PAM$^+$10c, PM13]. In this approach, Alice and Bob get the devices from a non-trusted provider Eve. The standard local-hidden-variable models described in 2.2.2 correspond to classical, essentially deterministic preparations in which the devices generate the measurement results given the choice of measurements, but independently of the input chosen by the other party. Bell inequalities are conditions satisfied by all these preparations, even when having access to all the measurement choices and results produced in previous steps [BCH$^+$02] (see 2.2.4 for more details). In turn, quantum correlations, obtained for example by measuring a maximally entangled two-qubit state with non-commuting measurements, can violate these inequalities. As we saw in 2.2.4, the violation of a Bell inequality witnesses the existence of non-local correlations and can be used by Alice and Bob to certify the quantum nature of their devices.

In what follows, it is shown how a classical Eve can mimic a Bell inequality violation when the measurement choices for Alice and Bob are performed following an algorithm, which is a standard practice in many Bell experiments to date. As above, it is not assumed that the algorithm is known by the eavesdropper. The result can be seen as a new loophole, named the *computability loophole*. For this loophole to apply, Eve has to make use of the inputs and outputs produced by the parties in previous steps [BCH$^+$02, PAM$^+$10c, PM13], as shown in Fig. 8.3.

The computability loophole is rather simple and works as follows: one of the devices, say Alice's, uses all the inputs chosen in previous steps by both parties to guess the next ones. For that, a time complexity class $\mathcal{C}$ is initially chosen by Eve. Since algorithms in such a class are computably enumerable, Alice's device will check all of them until it finds one that matches Alice's (or Bob's) bits given so far. A guess for the next inputs is done based on that algorithm (see Fig. 8.4 for a representation of the algorithm) and communicated to the

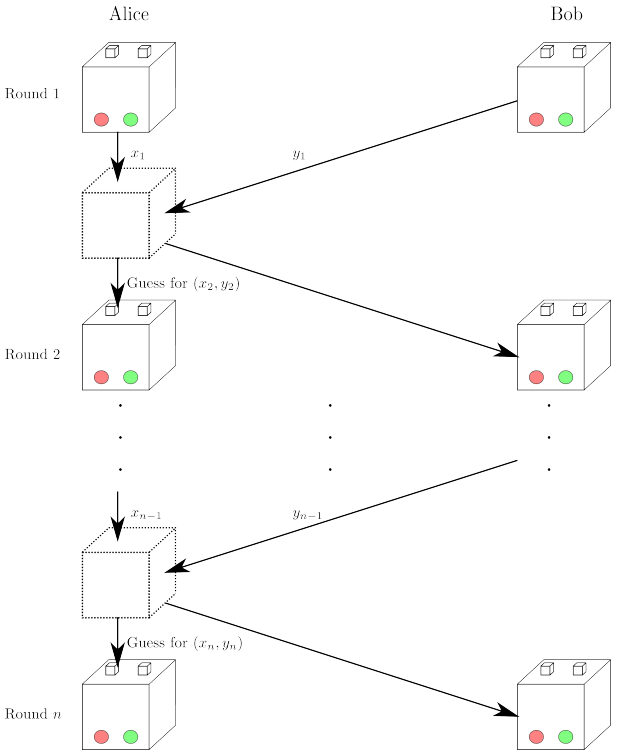

Figure 8.3.: Scheme for the Bell inequality computability loophole. After each round i, Alice's box receives the information about Bob's choice of measurement y_i. Using the information from all previous choices of inputs for both parties, Alice's box makes a prediction for what the inputs of the next round will be by using the presented algorithm. For simplicity in the notation, the attack is shown for a Bell test involving two measurements per party. As long as Alice, Bob or both use a computable sequence for their choice, the guess of the device will start being correct after a number of rounds. Once this happens, the boxes can simulate any probability distribution, both local and non-local. Therefore Alice and Bob can not rule out a classical eavesdropper having prepared their boxes.

other device before the next round of the Bell test is performed. Notice the boxes are not allowed to communicate during one round of the test as special relativity demands and is imposed by the no-signaling assumption, but they can communicate in-between rounds. If Eve's class $\mathcal{C}$ includes Alice's and/or Bob's algorithms, at some point the device will start guessing correctly and will

keep doing so forever. Of course, once the devices are able to guess the inputs of at least one of the parties, they can easily produce non-local correlations.

Seen bits:		1	0	1					
	s_1	0	0	0	0	0	0	0	...
	s_2	0	0	1	1	0	1	1	...
	s_3	1	0	0	0	1	0	1	...
	s_4	1	1	1	1	0	0	0	...
	s_5	0	1	0	1	0	1	0	...
First match:	s_6	1	0	1	0	1	1	1	... Next guess: 0
	s_7	1	1	0	1	1	0	1	...
	⋮								

Figure 8.4.: Algorithm to predict bits from previously seen ones. For simplicity it is presented for the case of 2 inputs. Alice's box enumerates all sequences s_i from the complexity class $\mathcal{C}$ and picks the first one whose first bits coincide with the ones seen. The guess Eve will make is the next bit from that sequence. If the actual sequence one wants to predict belongs to class $\mathcal{C}$, at some point, after seeing enough bits, all the predictions will be correct.

It should be noticed that since Alice's and Bob's algorithms belong to some time complexity class, they can never rule out such an eavesdropper. On the other hand, the eavesdropper, when choosing the class $\mathcal{C}$, is imposing how hard it is for Alice and Bob to avoid the loophole. The algorithm is shown as Algorithm 2, and for a more detailed description the reader is refered to the Appendix E.2.

Algorithm 2 A next-value algorithm for a time class $\mathcal{C}$ with computable bound t

Require: $n \in \mathbb{N}$

Ensure: $g(n)$, the next-value function for $\mathcal{C}$.

Let M_e be an enumeration of all Turing machines.

Let $n = \langle m_0, \ldots, m_{n-1} \rangle$ be the already seen bits from the sequence.

Let $\langle e, c \rangle$ be the least number such that

i. for $i \in \{0, \ldots, n\}$, $M_e(i)$ halts after at most $c \cdot t(|i|)$ many steps, where t is the computable time bound for class $\mathcal{C}$.

ii. for $i \in \{0, \ldots, n-1\}$ $M_e(i)$ outputs m_i

Output $M_e(n)$

Apart from being very demanding, a criticism to this loophole is that it only works in the long run, meaning that Alice and Bob will not see a fake violation

of a Bell inequality unless they run their experiment for long enough. But this brings the question of what's the validity of a violation that, in the long run, would have admited a local model. The only way to escape this loophole is by using a quantum random generator for the inputs, however, it is highly undesirable to depend on a non-local theory to test non-locality.

8.3. Discussion

The Church–Turing thesis is one of the most accepted postulates from computer science. As such, one can wonder what consequences would it have if it were, indeed, a law of Nature. In relation with quantum physics in this Chapter we showed how apparent paradoxes arise when we confront such thesis against the notion of what a proper mixed quantum state is and we introduced a new loophole for Bell tests.

First we showed that if Alice uses a computing device satisfying the Church-Turing thesis to prepare a seemingly proper mixture of $|0\rangle$ and $|1\rangle$ or a seemingly proper mixture of $|+\rangle$ and $|-\rangle$, both apparently yielding the maximally mixed state, Bob can distinguish both situations.

It is worth noticing that, although our algorithm halts in finite time, it can take extremely long, depending on the length of the shortest program that generates the needed prefix and the time it takes to find it. Nonetheless, what we have shown with this is that in both preparations, somehow, the resulting state has information about how it was prepared. Our algorithm can be thought of as a tool to prove that both preparations are indeed distinguishable, but there might be protocols that finish in shorter times. And even if there are not, the fact that both situations are distinguishable still holds, showing that having a computable preparation leaves a mark on the states it produces.

This apparent paradox can be easily resolved the following way: computable sequences have correlations that we are not taking into account. This means that Alice's choice is not given, as needed, by a set of independent and identically distributed random variables but by a computable sequence. The evident consequence of this is that Bob can distinguish both situations and our main results is to provide such an algorithm.

It is not clear whether Bob can associate a proper quantum state to each qubit leaving Alice's box. Let us imagine a situation in which Alice has already given Bob several qubits (as many qubits as Bob wanted to request from Alice), and we ask Bob to guess the next qubit —unknown to him—, with the sole promise that Alice, in the limit, will pick as many states from one bag as from the other (i.e. it is a balanced sequence). Since every prefix can be extended

to a computable sequence, no matter what Bob already knows about Alice's preparation, he cannot say anything about the next qubit. For instance, he can already know in what basis Alice is preparing each state (via the presented algorithm), and an extremely long prefix of the computable sequence that Alice is using. Still, he does not know if the next qubit will be $|0\rangle$ or $|1\rangle$ (if Alice prepares in the computational basis). The best description for that single qubit state that Bob can give, from the balanced sequence promise, is $\rho = \mathbb{I}/2$.

Interestingly, our results easily extend to other ensembles, and can for instance be applied to the mixed states experimentally produced using a classical random number generator of [AB09, LKPR10]. Our classical algorithm is also suitable for performing other seemingly impossible tasks. If Bob is presented with two states, one that is a proper computable mixture of the states $|0\rangle$ and $|1\rangle$ each with equal weight, and the other is an improper mixture yielding the maximally mixed state (for instance, one of the parts of a maximally entangled state), Bob can distinguish which is which by a slight modification of our algorithm. He just obtains two sequences, each by measuring σ_z to each state. Again, the problem is reduced to distinguishing, with high probability, a fair coin from a computable sequence, a task that we have already shown how to solve.

Then we showed that if either Alice or Bob choose their inputs for a Bell experiment in a computable way, an eavesdropper restricted to preparing deterministic devices can make them believe to have non-local boxes, thus creating a computability loophole. Notice that this scenario is equivalent to letting the boxes communicate before the runs and adapt accordingly, as is the case in the randomness expansion protocols [PM13, PAM$^+$10c] where our loophole would also apply if either Alice or Bob would use pseudo-randomness. There is no way of preventing this form of communication, unless some assumptions regarding the shielding of the devices are enforced.

From a fundamental perspective, our result answers the question about what type of randomness is necessary for having a valid violation of a Bell inequality: Alice and Bob's behaviour need to be non-algorithmic. Therefore no computable pseudo randomness criterion will suffice for a proper Bell inequality violation. It is natural to ask, at this point, where can Alice and Bob find sources of non-algorithmic randomness for their inputs. If they assume quantum mechanics, then flipping a perfectly modeled quantum coin would suffice (with probability one). However, as we mentioned it is not desirable to assume a non-local theory like quantum mechanics, in order to test non-locality.

9. Conclusions and outlook

In this thesis we have worked at the intersection between quantum foundations and quantum information technologies. Our work aimed at advancing our understanding of the reality of quantum mechanics and its place among all conceivable theories, and from that gained knowledge, propose novel information processing protocols.

The results achieved during the development of this thesis not only provided new insights into different topics such as derivations of quantum theory, randomness certification or the role of computability in quantum experiments, but also raised several new questions. In this Chapter, I will review the conclusions of this thesis and explore some of these new directions for future work.

Axiomatic reconstructions of quantum theory We have studied quantum theory in the broad space of generalized probabilistic theories with the aim of understanding what links its local structure with its global structure. We proved that, for the case of N qubits, the local structure of quantum theory is uniquely linked to its global structure through the reversibility of the dynamics. This in turn means that the viability of large scale quantum computing can be based on very concrete principles which can be experimentally tested at a local level without needing to test it in a macroscopic number of qubits simultaneously.

An interesting open problem of our work is extending our result to N systems of arbitrary dimension, i.e. going beyond the qubit case. The question here would be: is the combination of local quantum theory, local tomography and continuous reversibility still able to single out quantum theory? This would significantly strengthen our result. We conjecture that this is indeed the case.

A more ambitious open problem is whether just local tomography and the (not necessarily continuous) reversibility of the dynamics are enough to recover quantum theory. The kind of result we would hope for is that quantum theory is the *biggest* theory fulfilling both requirements and allowing for entanglement. We conjecture that all other theories fulfilling these principles could be seen as subtheories or restrictions on quantum theory.

Different results so far seem to indicate that this is a plausible possibility. First, in [MMPGA14] it was proved that, when considering two systems with local state spaces consisting on $d-$dimensional balls, no entangling dynamics

exist under the two mentioned principles: (continuous) reversibility and local tomography, unless $d = 3$ which is the quantum case.

A second observation supporting our conjecture is that, after about a decade trying to find interesting and consistent generalized probabilistic theories different from quantum theory, very few examples have been found. A remarkable example is the family of Generalized no-signaling theories we saw in Chapter 2. However, this family of theories has been shown to be highly problematic, i.e. having trivial reversible dynamics [GMCD10] and collapsing communication complexity [BBL$^+$06]. All this essentially means that it is not easy to find local state spaces that allow for rich entangling reversible dynamics. Given these facts, our conjecture is that the local state space of quantum theory is the biggest one admitting entangling reversible dynamics. This, together with our result and its possible extension to higher dimensional systems would mean that quantum theory can be derived from just two very general principles.

Device-independent randomness certification tasks We turned to the field of randomness certification and predictability in quantum theory and established results in three main projects.

Full randomness amplification: We studied what would be the biggest evidence of a completely unpredictable event in nature. Although the Bell theorem approaches the question from the right perspective, it requires initial perfect randomness to work thus falling into a circular argument. On the other hand, from no initial randomness, no final randomness can be certified on purely device-independent arguments. Recently, the mentioned question has been identified with an information-theoretic task called full randomness amplification. The aim of this task is to use a source of arbitrarily weakly random bits to produce a source of fully random bits, an impossible task using classical resources. Our contribution was to prove for the first time that full randomness amplification becomes possible using quantum nonlocality.

There are several interesting follow-up questions to our work. First, a limitation of our protocol is that it requires an infinite number of parties in order to achieve full amplification. Since our work was meant as a proof-of-principle, no optimization of resources was performed. As a result, the scaling on the number of devices demanded in our protocol makes it highly impractical. The question is then: can full randomness amplification be performed using a finite number of devices? A recent paper [BRG$^+$13] provided an answer to this question by constructing a full randomness amplification protocol that uses only 8 devices. Of course the devices need to be reused many times, but no assumptions on the memory of those devices is made.

Another follow-up question of our work concerns the class of arbitrarily weak randomness sources we show how to amplify. In Appendix B we show some relaxations on the class of sources we can amplify with our protocol i.e. not only the Santha-Vazirani sources. However, an interesting question appears: what is the most general source of randomness that can be amplified? This means pushing to its breaking point the initial motivation of our work, that is, what is the biggest evidence of a completely unpredictable event in nature?

Completeness of finite-size quantum events: We looked into a related question regarding the completeness of quantum theory for finite-size processes. We wondered whether there are processes as intrinsically random as quantum theory predicts under minimal assumptions. So far, all known processes were able to prove completeness of quantum theory in some asymptotic scenarios. This opened the interesting possibility that quantum mechanics could be an asymptotically correct approximation to some underlying theory. Our contribution was to provide a family of finite processes and prove they are as intrinsically random as they are observed to be. We proved this to hold under minimal assumptions: the validity of the no-signaling principle and the existence of arbitrarily weak randomness sources.

An interesting follow-up problem of our work is to find fully intrinsically random events that certify more than one bit of randomness. Are there fully intrinsically random processes that certify any amount of randomness? Another related problem concerns the scaling of our result. The family of processes we considered ends up certifying a perfectly random bit in the infinite number of parties limit. Analogously to our previous work, can we do the same with a finite number of parties? Finally, another open issue concerns the testing of our results in an experimental setting. As discussed in Chapter 5, given that experimental errors are unavoidable in a realistic setting, testing our results would require strenghtening the condition of 'arbitrarily weak initial randomness' to some small but finite amount characterized by δ in equation (5.1) of Chapter 5. The smaller the value of initial randomness δ the more stringent the test and the bigger the evidence and certainty about the completeness of quantum theory for those events.

Maximally nonlocal theories are not maximally random: We studied the relationship between the randomness capabilities of a theory and its degree of nonlocality. Our contribution was to show that maximally nonlocal theories cannot be maximally random. On the other hand, we show that quantum theory can certify maximal randomness in all dichotomic scenarios. Finally, we showed that quantum theory is not unique with respect to that capability i.e. there are theories allowing supra-quantum correlations which are nonetheless capable of certifying maximal randomness.

Our work opens several interesting venues for furture work. Concerning our first result, we gave an upper bound on the maximal amount of randomness that can be certified by assuming solely the no-signaling principle i.e. in maximally nonlocal theories. However, we know that, even if it has the correct scaling for a fixed number of outcomes and growing number of parties, the bound is not tight in general. It would be interesting to find the exact formula for the maximal amount of randomness that can be certified in every scenario. This could be seen as a fundamental relation linking randomness and the no-signaling principle. Concerning our second result, the open question is whether quantum theory can certify maximal randomness in general scenarios i.e. not only dichotomic. We conjecture that this is the case.

Another interesting question concerns how we defined maximal randomness. As is customary in randomness certification, we looked at the maximal probability of an adversary to guess the outcome of a Bell test for a given input. However, another reasonable definition could be one in which the average maximal probability over all inputs is considered. Studying how do maximally nonlocal theories behave with respect to quantum theory when such a definition is used would be an interesting question. Informally speaking, in our study we seemed to identify a tradeoff between the amount of randomness present across different inputs. That is, if one could certify a high amount of randomness in one input, other inputs would not contain so much randomness. These observations, however, deserve further work to be formalized.

Semi-device independent quantum information We focused on the nascent field of semi-device-independent quantum information noticing that future progress of the field was capped by the lack of an effective way to bound the strength of quantum correlations with local dimension constraints. Our contribution was to design a complete hierarchy of semidefinite programming relaxations to characterize those sets of quantum correlations. We explored several applications, the most notable of which was a device-independent test for high-dimensional entanglement.

As for the outlook of our work, even though our characterization is capable of treating relatively large scenarios and we proved its convergence, it is still computationally very demanding. Evidence of this could be seen in the different scenarios where we could not achieve tightness i.e. the upper bounds obtained by our method at the level of our hierarchy that a normal desktop could handle could not be matched by the lower bounds provided by the very efficient see-saw-type methods. In this context, a novel method has been developed in [NV14] based on noncommutative polynomial optimization theory which shows

an improvement over our method in terms of computational efficiency. Indeed, they could provide tightness in some scenarios where our method could not with a normal desktop.

As of next steps in the field of semi-device-independent quantum information, in my opinion we still need to identify novel applications where an assumption over the available quantum degrees of freedom (as the one required for semi-device-independent quantum information) is reasonable and provides a more significant advantage over the lack of such an assumption, that is, over the fully device-independent scenario.

Quantum randomness and computability In a nutshell, we studied what would be the implications of raising the Church-Turing thesis to the level of a postulate. In particular, the interaction of this new postulate with fundamental results of quantum theory. Our contribution comprises two main results. The first one concerns the preparation of ensembles by the computable mixture of different states. We show that mixing eigenstates of spin along one direction or an orthogonal one becomes distinguishable under the sole promise that the mixture was performed by computable means. Notice that if the mixture was performed using a fair coin instead of a computable process both methods would yield the maximally mixed state and would hence be undistinguishable.

The second result shows a loophole in Bell-like experiments. We show that if some of the parties in a Bell-like experiment use a computer to decide which measurements to make, then the computational resources of an eavesdropper have to be limited in order to have a proper observation of non-locality.

Our work opens a number of new and interesting research directions. Firstly, with this work we start a research programme that tries to find implications of computer science principles in physics. We gave here the first steps in that direction, but many other venues can be explored. One interesting question we are currently working on is whether Bohm-type theories can reproduce quantum predictions, such as Bell Inequality violations, even when restricted to computable means. It is clear that, if a Bohm-type theory is to reproduce a Bell Inequality violation, it needs to use hidden signaling. Our initial results seem to indicate that, if only computable resources are available to the Bohm-type theory, there exist algorithms that can extract that hidden signaling, exploiting it to send messages. The implications of this work will be explored in a forthcoming paper.

There is another interesting question concerning our first result. We provided an algorithm that was proven to distinguish two situations with arbitrarily high probability and in finite time. However, there is an interesting question we did

not explore in our work: how long would our algorithm typically take in some practical situations? For instance, it would be interesting to look at the halting time of a modified algorithm suited for a simpler scenario. For example, let us imagine that we know that the randomness used to computably prepare quantum states is obtained through the default random number generator function of a particular programming language, such as MATLAB. Imagine however that we do not know what seed was used in the process. What would be a typical amount of time needed to distinguish the two situations with, say, 90% probability of success?

A. Proofs of convergence of chapter 7

A.1. Convergence of expanded bodies

In general, a positive semidefinite operator $\Gamma^{(n)}$ of the form (7.2.2) will not possess a **moment representation**, i.e., there will not exist a state ρ and projector operators satisfying

$$E_b^y E_{b'}^y = F_c^z F_{c'}^z = 0, \tag{A.0}$$

for $b \neq b', c \neq c'$, and

$$[E_c^y, F_c^z] = 0, \tag{A.0}$$

such that $c_s^{k,j} \equiv \mathrm{tr}(|j\rangle\langle k| \otimes s\rho)$.

One can, however, prove the following result.

Theorem A.1. *Let $\Gamma^{(n)}$ be a positive semidefinite matrix of the form (7.2.2). Then, there exist a finite-dimensional Hilbert space $\mathcal{H}$, a normalized state $\rho \in B(\mathbb{C}^d \otimes \mathcal{H})$ and projector operators $\{\hat{E}_b^y, \hat{F}_c^z\} \subset B(\mathcal{H})$ satisfying (A.1) such that*

$$c_s^{k,j} = tr\{\rho(|j\rangle\langle k| \otimes \hat{s})\}, \tag{A.0}$$

for any sequence $\hat{s}$ of the operators $\{\hat{E}_b^y, \hat{F}_c^z\}$ with $|s| \leq 2n$.

Note that, due to the structure of the coefficients $\{c_s^{k,j}\}$, even though the commutator $[E_c^y, F_c^z]$ may be different from zero, the identity

$$\mathrm{tr}\{\rho(|j\rangle\langle k| \otimes s\hat{E}_b^y \hat{F}_c^z \tilde{s})\} = \mathrm{tr}\{\rho(|j\rangle\langle k| \otimes s\hat{F}_c^z \hat{E}_b^y \tilde{s})\} \tag{A.0}$$

must hold as long as $|sE_b^y F_c^z \tilde{s}| \leq 2n$, since both operator products are associated to the same 'logical' sequence.

Also notice that, if only one party, say Bob, was expanded, eq. (A.1) implies that we achieve convergence with $n = 1$.

Proof. The condition $\Gamma^{(n)} \geq 0$ implies [HJ12] that

$$c_{t^\dagger s}^{k,j} = \Gamma^{(n)}_{(k,s),(j,t)} = \langle \psi_t^j | \psi_s^k \rangle, \tag{A.0}$$

for some collection of vectors $\{|\psi_s^k\rangle\}$. Here, as in Chapter 7, the variables s, t are used to represent operator products; k, j, natural numbers ranging from 1 to d. With a slight abuse of notation, if s denotes a null sequence, the corresponding coefficient $c_s^{k,j}$ will be taken equal to zero.

Now, define the vector

$$|\phi\rangle \equiv \sum_k |k\rangle \left|\psi_{\mathbb{I}}^k\right\rangle. \tag{A.0}$$

It is immediate that this vector is normalized. Indeed, note that

$$\langle\phi|\phi\rangle = \sum_k \langle\psi_{\mathbb{I}}^k|\psi_{\mathbb{I}}^k\rangle = \sum_k c_{\mathbb{I}}^{k,k} = 1. \tag{A.0}$$

We will hence identify $|\phi\rangle$ with the normalized state in the theorem, i.e., $\rho = |\phi\rangle\langle\phi|$.

Now, define the subspaces

$$\begin{aligned}\mathcal{H}_b^y &= \text{span}\{|\psi_s^k\rangle : s = E_b^y \tilde{s}, k = 0, ..., d-1\},\\ \mathcal{H}_c^z &= \text{span}\{|\psi_s^k\rangle : s = F_c^z \tilde{s}, k = 0, ..., d-1\}.\end{aligned} \tag{A.0}$$

For $b \neq b'$, The fact that $0 = c_0^{j,k} = \langle\psi^j_{E_{b'}^y s}|\psi^k_{E_b^y \tilde{s}}\rangle$ implies that $\mathcal{H}_b^y \perp \mathcal{H}_{b'}^y$, and likewise we have that $\mathcal{H}_c^z \perp \mathcal{H}_{c'}^z$, for $c \neq c'$. It follows that the projectors

$$\hat{E}_b^y \equiv \text{proj}(\mathcal{H}_b^y), \hat{F}_c^z \equiv \text{proj}(\mathcal{H}_c^z) \tag{A.0}$$

satisfy

$$\hat{E}_b^y \hat{E}_{b'}^y = \delta_{bb'} \hat{E}_b^y, \hat{F}_c^z \hat{F}_{c'}^z = \delta_{cc'} \hat{F}_c^z. \tag{A.0}$$

Let us explore how these operators act over the vectors $\{|\psi_s^k\rangle\}$. We have that

$$\hat{E}_b^y \left|\psi_s^k\right\rangle = \hat{E}_b^y \sum_{b' \neq \tilde{b}} \left|\psi^k_{E_{b'}^y s}\right\rangle + \hat{E}_b^y |\text{rest}\rangle \tag{A.0}$$

with

$$|\text{rest}\rangle = \left|\psi_s^k\right\rangle - \sum_{b' \neq \tilde{b}} \left|\psi^k_{E_{b'}^y s}\right\rangle \tag{A.0}$$

(we remind the reader that $\tilde{b}$ represents the measurement outcome not included in the expansion of Bob's body).

Due to the orthogonality relations $\mathcal{H}_b^y \perp \mathcal{H}_{b'}^y$, for $b \neq b'$, the first term on the right hand side of eq. (A.1) is $\left|\psi^k_{E_b^y s}\right\rangle$. As for the second term, notice that

$$\langle \psi^j_{E_b^y t} | \text{rest} \rangle = c^{j,k}_{t^\dagger E_b^y s} - c^{j,k}_{t^\dagger E_b^y s} = 0. \tag{A.0}$$

It follows that $\hat{E}_b^y \left|\text{rest}\right\rangle = 0$. Putting all together, we have that

$$\hat{E}_b^y \left|\psi_s^k\right\rangle = \left|\psi^k_{E_b^y s}\right\rangle, \tag{A.0}$$

and, similarly,

$$\hat{F}_c^z \left|\psi_s^k\right\rangle = \left|\psi^k_{F_c^z s}\right\rangle. \tag{A.0}$$

It follows by induction that, for any sequence $\hat{s}$ of the operators $\{\hat{E}_b^y, \hat{F}_c^z\}$,

$$\hat{s} \left|\psi_{\mathbb{I}}^k\right\rangle = \left|\psi_s^k\right\rangle. \tag{A.0}$$

Finally, we arrive at

$$\begin{aligned} \text{tr}\{\rho(|j\rangle\langle k| \otimes \hat{t}^\dagger \hat{s})\} \quad &= \left\langle \psi_{\mathbb{I}}^j \right| \hat{t}^\dagger \hat{s} \left|\psi_{\mathbb{I}}^k\right\rangle = \\ &= \langle \psi_t^j | \psi_s^k \rangle = c^{j,k}_{t^\dagger s}. \end{aligned} \tag{A.0}$$

□

Following the lines of [NPA08], the convergence of the scheme follows from the fact that, for any sequence of positive semidefinite moment matrices $(\Gamma^{(n)})_n$ such that (7.2.2) holds, there exists a set of vectors $\{\left|\psi_s^k\right\rangle : k = 0, ..., d-1\} \subset \mathcal{H}$ which allow (using the same construction as in the previous theorem) to build projector operators $\{E_b^y, F_c^z\} \subset B(\mathcal{H})$ and a quantum state $\rho \in B(\mathbb{C}^d \otimes \mathcal{H})$ which satisfy eq. (7.2.2). The proof is nearly identical to the one in [NPA08], and so it will not be included in this Appendix.

A.2. Convergence of heads, legs and extended bodies

The purpose of this Appendix is to prove that, for very long legs, the matrix that results when we trace out from $\Gamma^{(n)}$ all circles but one on each of the L legs, the result can be approximated by an expression of the form

$$\sum_k p_k \bigotimes_{l=1}^{L} \left|u_l^k\right\rangle\left\langle u_l^k\right| \otimes \tilde{\Gamma}_k^{(n)}, \tag{A.0}$$

where $p_k \geq 0$, $\sum p_k = 1$ and $\tilde{\Gamma}_k^{(n)}$ is a generalized moment matrix representing Alice's head and Bob and Charlie's expanded bodies. Also, in the above expression, orthogonal legs remain orthogonal. In combination with the results of the previous Appendix, this will show that, taking the limits $\lim_{n\to\infty}(\lim_{N\to\infty})$, the proposed hierarchy achieves convergence.

First, denoting by $\mathcal{H}_L$ the Hilbert space associated to Alice's legs, note that, for any positive semidefinite operator $M \in B(\mathcal{H}_L)$,

$$\mathrm{tr}_L\{(M_L \otimes \mathbb{I})\Gamma^{(n)}\} \tag{A.0}$$

is a positive semidefinite operator of the form (7.2.2), but not necessarily fulfilling the normalization condition (7.2.2).

Now, given the symmetric space of $\mathbb{C}^d$, $\mathcal{H}_d^N$, consider the trace-preserving CP map $\Lambda : B(\mathcal{H}_d^N) \to B(\mathbb{C}^d)$ defined by:

$$\Lambda(\bullet) \equiv \begin{pmatrix} N+d-1 \\ N \end{pmatrix} \int \mathrm{tr}(|\phi\rangle\langle\phi|^N \bullet)\, |\phi\rangle\langle\phi|\, d\phi. \tag{A.0}$$

This map was proposed in [NOP09] to study the convergence of the DPS hierarchy [DPS05]. In [NOP09], it was shown that it is equivalent to the partially depolarizing channel:

$$\Lambda(\bullet) \equiv \frac{N}{N+d}\mathrm{tr}_{N-1}(\bullet) + \frac{d}{N+d}\mathrm{tr}(\bullet)\mathbb{I}_d. \tag{A.0}$$

By (A.2) it is clear that, applying the map Λ to any leg in $\Gamma^{(n)}$, the resulting matrix $\hat{\Gamma}$ is of the form (A.2). Orthogonal legs may not remain orthogonal, though. However, for any two orthogonal legs C, D, by formula (A.2), in the limit of $N \gg 1$, $\mathrm{tr}(V(C,D)\hat{\Gamma})$ tends to zero, thus guaranteeing asymptotic orthogonality. Finally, also by eq. (A.2), $\hat{\Gamma}$ can be made arbitrarily close in trace norm to the partial trace of $\Gamma^{(n)}$. Note also that the speed of convergence does not depend on the Hilbert space dimension of the expanded bodies, but on the total trace of $\Gamma^{(n)}$.

Finally, let us remark that maps of the form (A.2) converge to the identity channel as $O(d/N)$. In order to improve this rate, one can use a second, more complicated map described also in [NOP09], which induces an $O((d/N)^2)$ convergence. Beware, though! Such a map can only be applied when the PPT condition has been enforced on $\lceil N/2 \rceil$ circles of each of the legs [NOP09].

B. Proof of full randomness amplification

B.1. Preliminaries

B.1.1. Notation

Before entering the details of the Theorem in chapter 4, let us introduce a convenient notation. In what follows, we sometimes treat conditional probability distributions as vectors. To avoid ambiguities, we explicitly label the vectors describing probability distributions with the arguments of the distributions in upper case. Thus, for example, we denote by $P(\mathbf{A}|\mathbf{X})$ the $(2^5 \times 2^5)$-dimensional vector with components $P(\mathbf{a}|\mathbf{x})$ for all $\mathbf{a}, \mathbf{x} \in \{0,1\}^5$. With this notation, the five-partite Mermin inequality can be written as the scalar product

$$I \cdot P(\mathbf{A}|\mathbf{X}) = \sum_{\mathbf{a},\mathbf{x}} I(\mathbf{a},\mathbf{x}) P(\mathbf{a}|\mathbf{x}) \geq 6 \ . \tag{B.0}$$

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

$$I^{\otimes N_d} \cdot P(B|Y) = \sum_{\mathbf{a}_1,\ldots\mathbf{a}_{N_d}} \sum_{\mathbf{x}_1,\ldots\mathbf{x}_{N_d}} \left[\prod_{i=1}^{N_d} I(\mathbf{a}_i,\mathbf{x}_i) \right] P(\mathbf{a}_1,\ldots\mathbf{a}_{N_d}|\mathbf{x}_1,\ldots\mathbf{x}_{N_d}) \ . \tag{B.0}$$

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

B.1.2. The ϵ-source

Consider the probability $P\left(x_j|x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n,e\right)$. It represents how the bits produced by the source $\mathcal{S}$ behave ie. such that the probability that bit j takes a given value x_j, conditioned on all the other bits produced by the source as well as on any variable e outside of the future light-cone of all x_j's, is

bounded by

$$\epsilon \leq P\left(x_j | x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n, e\right) \leq 1 - \epsilon, \tag{B.0}$$

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

$$\epsilon^n \leq P\left(x_1, \ldots, x_n | e\right) \leq (1 - \epsilon)^n. \tag{B.0}$$

This bound plays a key role in the estimation part of our protocol.

B.2. Proofs

The proof of the Theorem in chapter 4 relies on two crucial lemmas, which are stated and proven in Sections B.2.2 and B.2.3, respectively. The first lemma bounds the distinguishability between the distribution distilled from a block of N_d quintuplets and the ideal free random bit as function of the Bell violation (B.1.1) in each quintuplet. In particular, it guarantees that, if the correlations of all quintuplets in a given block violate inequality (B.1.1) sufficiently much, the bit distilled from the block will be indistinguishable from an ideal free random bit. The second lemma is required to guarantee that, if the statistics observed in all blocks but the distilling one are consistent with a maximal violation of inequality (B.1.1), the violation of the distilling block will be arbitrarily large.

B.2.1. Proof of the Theorem

We begin with the identity

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

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

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

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

From the definition of the guessing probability, we have

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

where the inequality is due to Lemma B.1 in Section B.2.2, we have used the no-signaling condition through $P(\tilde{y},t|z,g=1)=P(\tilde{y},t|g=1)$, in the second equality, and Bayes rule in the second and sixth equalities. From (B.-6) and Lemma B.2 in Section B.2.3, we obtain

$$
P(\text{guess}|g=1) \;\leq\; \frac{1}{2}+\frac{3\sqrt{N_d}}{2}\left[\alpha^{N_d}+\frac{2\,N_b^{\log_2(1-\epsilon)}}{P(g=1)}\left(32\beta\epsilon^{-5}\right)^{N_d}\right]. \qquad \text{(B.-6)}
$$

Finally, substituting bound (B.2.1) and equality (B.2.1) into (B.2.1), we obtain

$$
P(\text{guess}) \;\leq\; \frac{1}{2}+\frac{3\sqrt{N_d}}{2}\left[P(g=1)\,\alpha^{N_d}+2\,N_b^{\log_2(1-\epsilon)}\left(32\beta\epsilon^{-5}\right)^{N_d}\right], \quad \text{(B.-6)}
$$

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

$$
P(\text{guess}) \;\leq\; \frac{1}{2}+\frac{3\sqrt{N_d}}{2}\left[\alpha^{N_d}+2\,N_b^{\log_2(1-\epsilon)}\left(32\beta\epsilon^{-5}\right)^{N_d}\right]. \qquad \text{(B.-6)}
$$

and in turn, proves the Theorem in chapter 4.

B.2.2. Statement and proof of Lemma B.1

As mentioned, Lemma B.1 provides a bound on the distinguishability between the probability distribution obtained after distilling a block of N_d quintuplets and an ideal free random bit in terms of the Bell violation (B.1.1) in each quintuplet. The proof of Lemma B.1, in turn, requires two more lemmas, Lemma B.3 and Lemma B.4, stated and proven in Section B.2.4.

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

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

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

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

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

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

$$P(\mathbf{w}|y) = \left(\bigotimes_{i=1}^{N_d} \Gamma_{w_i}^{\mathbf{x}_i} \right) \cdot P(B|Y). \quad \text{(B.-6)}$$

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

$$\begin{aligned}
&\sum_k \max_z \sum_e \left| P(k,e|y,z) - \frac{1}{2} P(e|y,z) \right| \\
\leq\ & \sum_k \max_z \sum_e P(e|z) \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right| \cdot P(B|Y,e,z) \\
=\ & \sum_k \max_z \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right| \cdot \left(\sum_e P(e|z) P(B|Y,e,z) \right) \\
=\ & \sum_k \left| \sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2} \right) \bigotimes_{i=1}^{N_d} \Gamma^{\mathbf{x}_i}_{w_i} \right| \cdot P(B|Y), \qquad \text{(B.-8)}
\end{aligned}$$

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

Take $\Lambda^{\mathbf{x}_i}_w$ to be equal to the vectors $\Lambda^{\mathbf{x}_0}_w$ in Lemma B.3. These vectors then satisfy the bounds (B.-17) and (B.4) in the same Lemma. Take f to be equal to the function whose existence is proven in Lemma B.4. Note that the conditions needed for this Lemma to apply are satisfied because of bound (B.-17) in Lemma B.3, and because the free parameter $N_d \geq 130$ satisfies $\left(3\sqrt{N_d}\right)^{-1/N_d} \geq \gamma = 0.9732$. With this choice of f and $\Lambda^{\mathbf{x}_i}_w$, bound (B.-8) becomes

$$\begin{aligned}
&\sum_k \max_z \sum_e \left| P(k,e|y,z) - \frac{1}{2} P(e|y,z) \right| \\
\leq\ & \sum_k 3\sqrt{N_d} \left(\bigotimes_{i=1}^{N_d} \Omega^{\mathbf{x}_i} \right) \cdot P(B|Y) \\
\leq\ & 6\sqrt{N_d} \left(\alpha C + \beta I \right)^{\otimes N_d} \cdot P(B|Y)\ , \qquad \text{(B.-9)}
\end{aligned}$$

where we have used $\Omega^{\mathbf{x}_i} = \sqrt{(\Gamma^{\mathbf{x}_i}_0)^2 + (\Gamma^{\mathbf{x}_i}_1)^2}$, $\sum_k 3 = 6$, bound (B.-17) in Lemma B.3 and bound (B.4) in Lemma B.4. □

B.2.3. Statement and proof of Lemma B.2

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

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

$$(\alpha C+\beta I)^{\otimes N_d}\cdot P(\tilde{B}|\tilde{Y},g=1) \;\leq\; \alpha^{N_d}+\frac{2\,N_b^{\log_2(1-\epsilon)}}{P(g=1)}\left(32\beta\epsilon^{-5}\right)^{N_d}\;. \quad \text{(B.-9)}$$

Proof of Lemma B.2. According to definition (see chapter 4)

$$r[b,y]=\begin{cases}1 & \text{if } I(\mathbf{a}_1,\mathbf{x}_1)=\cdots=I(\mathbf{a}_{N_d},\mathbf{x}_{N_d})=0\\ 0 & \text{otherwise}\end{cases} \quad \text{(B.-9)}$$

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

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

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

$$\begin{aligned}
& (\alpha C + \beta I)^{\otimes N_d} \cdot P(B|Y, g=1) \\
= & \sum_{\mathbf{a}_1,\ldots\mathbf{a}_{N_d}} \sum_{\mathbf{x}_1,\ldots\mathbf{x}_{N_d}} \prod_{i=1}^{N_d} \left[\alpha\, 2^{-5} + \beta I(\mathbf{a}_i, \mathbf{x}_i)\right] P(\mathbf{a}_1, \ldots \mathbf{a}_{N_d} | \mathbf{x}_1, \ldots \mathbf{x}_{N_d}, g=1) \\
\leq & \sum_{b,y} \left[\left(\alpha\, 2^{-5}\right)^{N_d} + (2\beta)^{N_d} \delta^0_{r[b,y]}\right] P(b|y, g=1) \\
= & \alpha^{N_d} \sum_y 2^{-5N_d} + (2\beta)^{N_d} \sum_y P(r=0|y, g=1) \\
= & \alpha^{N_d} + (2\beta)^{N_d} \sum_y P(r=0|y, g=1) \\
= & \alpha^{N_d} + (2\beta)^{N_d} \sum_y \frac{P(r=0, y|g=1)}{P(y|g=1)} \ . \qquad \text{(B.-16)}
\end{aligned}$$

We can now bound $P(y|g=1)$ taking into account that y denotes a $5N_d$-bit string generated by the ϵ-source $\mathcal{S}$ that remains after step 2 in the protocol. Note that only half of the 32 possible 5-bit inputs $\mathbf{x}$ generated by the source belong to $\mathcal{X}$ and remain after step 2. Thus, $P((\mathbf{x}_1, \ldots, \mathbf{x}_{N_d}) \in \mathcal{X}^{N_d} | g=1) \leq 16^{N_d}(1-\epsilon)^{5N_d}$, where we used (B.1.2). This, together with $P((\mathbf{x}_1, \ldots, \mathbf{x}_{N_d})|g=1) \geq \epsilon^{5N_d}$ implies that

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

Substituting this bound in (B.-16), and summing over y, gives

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

In what follows we use the notation

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

According to step 4 in the protocol described in chapter 4, it aborts ($g=0$) if there is at least a "not right" block ($r[b_j, y_j] = 0$ for some $j \neq l$). While abortion also happens if there are more than one "not right" block, in what follows we lower-bound $P(g=0)$ by the probability that there is only one "not

right" block:

$$\begin{aligned}
1 &\geq P(g=0) \\
&\geq \sum_{l=1}^{N_b} P(l) \sum_{l'=1,\, l'\neq l}^{N_b} P(1_1, \ldots 1_{l-1}, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&\geq \sum_{l} P(l) \sum_{l'\neq l} P(1_1, \ldots 1_{l-1}, 1_l, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&= \sum_{l'} \left[\textstyle\sum_{l\neq l'} P(l)\right] P(1_1, \ldots 1_{l-1}, 1_l, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&= \sum_{l'} [1 - P(l')]\, P(1_1, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}), \qquad \text{(B.-19)}
\end{aligned}$$

where, when performing the sum over l, we have used that $P(1_1, \ldots 1_{l-1}, 1_l, 1_{l+1}, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \equiv P(1_1, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b})$ does not depend on l. Bound (B.1.2) implies

$$\frac{1-P(l)}{P(l)} \;\geq\; \frac{1-(1-\epsilon)^{\log_2 N_b}}{(1-\epsilon)^{\log_2 N_b}} \;=\; N_b^{\log_2 \frac{1}{1-\epsilon}} - 1 \;\geq\; \frac{N_b^{\log_2 \frac{1}{1-\epsilon}}}{2}\,, \qquad \text{(B.-19)}$$

where the last inequality holds for sufficiently large N_b. Using this and (B.-19), we obtain

$$\begin{aligned}
1 &\geq \frac{1}{2} \sum_{l'} N_b^{\log_2 \frac{1}{1-\epsilon}}\, P(l')\, P(1_1, \ldots 1_{l'-1}, 0_{l'}, 1_{l'+1}, \ldots 1_{N_b}) \\
&\geq \frac{1}{2}\, N_b^{\log_2 \frac{1}{1-\epsilon}}\, P(\tilde{r}=0, g=1)\,, \qquad \text{(B.-19)}
\end{aligned}$$

where $\tilde{r} = r[b_l, y_l]$. This together with (B.2.3) implies

$$\begin{aligned}
& (\alpha C + \beta I)^{\otimes N_d} \cdot P(\tilde{B}|\tilde{Y}, g=1) \\
\leq\; & \alpha^{N_d} + (2\beta)^{N_d} \left(\frac{16(1-\epsilon)^5}{\epsilon^5}\right)^{N_d} P(\tilde{r}=0|g=1) \qquad \text{(B.-19)} \\
\leq\; & \alpha^{N_d} + \frac{2}{P(g=1)} \left(\frac{32\beta(1-\epsilon)^5}{\epsilon^5}\right)^{N_d} N_b^{\log_2(1-\epsilon)}\,, \qquad \text{(B.-18)}
\end{aligned}$$

where, in the second inequality, Bayes rule was again invoked. Inequality (B.-18), in turn, implies (B.2). □

B.2.4. Statement and proof of the additional Lemmas

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

$$\begin{aligned} & \sqrt{\left[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right]^2 + \left[M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right]^2} \\ \leq \; & \alpha C(\mathbf{a},\mathbf{x}) + \beta I(\mathbf{a},\mathbf{x}) + \Lambda_2^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) \end{aligned} \tag{B.-18}$$

and

$$\begin{aligned} & \left|M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right| \\ \leq \; & \gamma \sqrt{\left[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right]^2 + \left[M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right]^2} \end{aligned} \tag{B.-18}$$

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

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

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

$$D \cdot \Lambda_w^{\mathbf{x}_0} = \mathbf{0} \tag{B.-18}$$

for $w \in \{0,1,2\}$, where $\mathbf{0}$ is the zero vector and D is a matrix whose rows constitute a basis of non-signaling probability distributions. A geometrical interpretation of constraint (B.-17) is that the point in the plane with coordinates $[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}), M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})] \in \mathbb{R}^2$ is inside a circle of radius $\alpha C(\mathbf{a},\mathbf{x}) + \beta I(\mathbf{a},\mathbf{x}) + \Lambda_2^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})$ centered at the origin. All points inside an octagon inscribed in this circle also satisfy constraint (B.-17). The points of such an inscribed octagon are the ones satisfying the following set of linear constraints:

$$\begin{aligned} & \left[M_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_0^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right] \eta \cos\theta + \left[M_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) + \Lambda_1^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right] \eta \sin\theta \\ \leq \; & \alpha C(\mathbf{a},\mathbf{x}) + \beta I(\mathbf{a},\mathbf{x}) + \Lambda_2^{\mathbf{x}_0}(\mathbf{a},\mathbf{x}) \; , \end{aligned} \tag{B.-18}$$

for all $\theta \in \{\frac{\pi}{8}, \frac{3\pi}{8}, \frac{5\pi}{8}, \frac{7\pi}{8}, \frac{9\pi}{8}, \frac{11\pi}{8}, \frac{13\pi}{8}, \frac{15\pi}{8}\}$, where $\eta = (\cos\frac{\pi}{8})^{-1} \approx 1.082$. In other words, the eight conditions (B.-17) imply constraint (B.-17). From now on, we only consider these eight linear constraints (B.-17). With a bit of algebra, one can see that inequality (B.-17) is equivalent to the two almost linear

inequalities there was an error in the following equation, as the pre-factor in terms of γ was wrong. Please check what was computed and how it affects to γ and, then, to the value of N_d

$$\pm\left[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})+\Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right] \;\leq\; \sqrt{\frac{\gamma^2}{1-\gamma^2}}\left|M_{\overline{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})+\Lambda_{\overline{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right| \;, \qquad \text{(B.-18)}$$

for all $w\in\{0,1\}$, where $\overline{w}=1-w$. Clearly, the problem is not linear because of the absolute values. The computation described in what follows constitutes a trick to make a good guess for the signs of the terms in the absolute value of (B.2.4), so that the problem can be made linear by adding extra constraints.

The first computational step consists of a linear-programming minimization of α subject to the constraints (B.2.4), (B.-17), where the minimization is performed over the variables $\alpha,\beta,\Lambda_0^{\mathbf{x}_0},\Lambda_1^{\mathbf{x}_0},\Lambda_2^{\mathbf{x}_0}$. This step serves to guess the signs

$$\sigma_w(\mathbf{a},\mathbf{x}) \;=\; \mathrm{sign}[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})+\Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})] \;, \qquad \text{(B.-18)}$$

for all $w,\mathbf{a},\mathbf{x}$, where the value of $\Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})$ corresponds to the solution of the above minimization. Once we have identified all these signs, we can write the inequalities (B.2.4) in a linear fashion:

$$\sigma_w(\mathbf{a},\mathbf{x})\left[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})+\Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right]\geq 0 \;,$$
$$\sigma_w(\mathbf{a},\mathbf{x})\left[M_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})+\Lambda_w^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right]\leq \sqrt{\frac{\gamma^2}{1-\gamma^2}}\,\sigma_{\overline{w}}(\mathbf{a},\mathbf{x})\left[M_{\overline{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})+\Lambda_{\overline{w}}^{\mathbf{x}_0}(\mathbf{a},\mathbf{x})\right] \;,$$

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

The second computational step consists of a linear-programming minimization of α subjected to the constraints (B.2.4), (B.-17), (B.2.4), (B.2.4), over the variables $\alpha,\beta,\Lambda_0^{\mathbf{x}_0},\Lambda_1^{\mathbf{x}_0},\Lambda_2^{\mathbf{x}_0}$. Clearly, any solution to this problem is also a solution to the original formulation of the Lemma. The minimization was performed for any $\mathbf{x}_0\in\mathcal{X}$ and the values of α,β turned out to be independent of $\mathbf{x}_0\in\mathcal{X}$. These obtained numerical values are the ones appearing in the formulation of the Lemma. □

Note that Lemma B.3 allows one to bound the predictability of $\mathrm{maj}(\mathbf{a})$ by a linear function of the 5-party Mermin violation. This can be seen by computing $\Gamma_w^{\mathbf{x}_0}\cdot P(\mathbf{A}|\mathbf{X})$ and applying the bounds in the Lemma. In principle, one expects this bound to exist, as the predictability is smaller than one at the point of maximal violation, as proven in lemma 4.1 in chapter 4, and equal to one at

the point of no violation. However, we were unable to find it. This is why we had to resort to the linear optimization technique given above, which moreover provides the bounds (B.-17) and (B.-17) necessary for the security proof.

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

$$\left|\Gamma^i_w(\mathbf{a},\mathbf{x})\right| \leq \left(3\sqrt{N_d}\right)^{-1/N_d} \Omega_i(\mathbf{a},\mathbf{x}) \ , \tag{B.-18}$$

where $\Omega_i(\mathbf{a},\mathbf{x}) = \sqrt{\Gamma^i_0(\mathbf{a},\mathbf{x})^2 + \Gamma^i_1(\mathbf{a},\mathbf{x})^2}$. There exists a function $f : \{0,1\}^{N_d} \to \{0,1\}$ such that for each sequence $(\mathbf{a}_1,\mathbf{x}_1),\dots(\mathbf{a}_{N_d},\mathbf{x}_{N_d})$ and any $k \in \{0,1\}$ we have

$$\left|\sum_{\mathbf{w}} \left(\delta^k_{f(\mathbf{w})} - \frac{1}{2}\right) \prod_{i=1}^{N_d} \Gamma^i_{w_i}(\mathbf{a}_i,\mathbf{x}_i)\right| \ \leq \ 3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i(\mathbf{a}_i,\mathbf{x}_i) \ , \tag{B.-18}$$

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

To get some intuition on why the bound (B.4) holds, it is useful to interpret $\mathbf{w}$ as a label of the 2^{N_d} steps in a random walk. For a fixed function $f(\mathbf{w})$, the variable $(\delta^k_{f(\mathbf{w})} - \frac{1}{2}) = \pm\frac{1}{2}$ specifies one trajectory of the random walk, and the number $\prod_{i=1}^{N_d} \Gamma^i_{w_i}(\mathbf{a}_i,\mathbf{x}_i)$ can be interpreted as the length of the step $\mathbf{w}$. It is expected that for a sufficiently generic trajectory, the final position of the particle will be of the order of its standard deviation

$$\sqrt{\sum_{\mathbf{w}} \left(\frac{1}{2}\prod_{i=1}^{N_d} \Gamma^i_{w_i}(\mathbf{a}_i,\mathbf{x}_i)\right)^2} = \sqrt{\frac{1}{4}\prod_{i=1}^{N_d}\sum_{w_i} \Gamma^i_{w_i}(\mathbf{a}_i,\mathbf{x}_i)^2} = \frac{1}{2}\prod_{i=1}^{N_d} \Omega_i(\mathbf{a}_i,\mathbf{x}_i) \ ,$$

which is essentially what we find in (B.4). In what follows, we provide a rigorous proof.

Proof of Lemma B.4. First, note that for a sequence $(\mathbf{a}_1,\mathbf{x}_1),\dots(\mathbf{a}_{N_d},\mathbf{x}_{N_d})$ for which there is at least one value of $i \in \{1,\dots N_d\}$ satisfying $\Gamma^i_0(\mathbf{a}_i,\mathbf{x}_i) = \Gamma^i_1(\mathbf{a}_i,\mathbf{x}_i) = 0$, both the left-hand side and the right-hand side of (B.4) are equal to zero, hence, inequality (B.4) is satisfied independently of the function f. Therefore, in what follows, we only consider sequences $(\mathbf{a}_1,\mathbf{x}_1),\dots(\mathbf{a}_{N_d},\mathbf{x}_{N_d})$ for which either $\Gamma^i_0(\mathbf{a}_i,\mathbf{x}_i) \neq 0$ or $\Gamma^i_1(\mathbf{a}_i,\mathbf{x}_i) \neq 0$, for all $i = 1,\dots N_d$. Or, equivalently, we consider sequences such that

$$\prod_{i=1}^{N_d} \Omega_i(\mathbf{a}_i,\mathbf{x}_i) > 0 \ . \tag{B.-18}$$

The existence of the function f satisfying (B.4) for all such sequences is shown with a probabilistic argument. We consider the situation where f is picked from the set of all functions mapping $\{0,1\}^{N_d}$ to $\{0,1\}$ with uniform probability, and upper-bound the probability that the chosen function does not satisfy the constraint (B.4) for all k and all sequences $(\mathbf{a}_1, \mathbf{x}_1), \dots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ satisfying (B.2.4). This upper bound is shown to be smaller than one. Therefore there must exist at least one function satisfying (B.4).

For each $\mathbf{w} \in \{0,1\}^{N_d}$ consider the random variable $F_{\mathbf{w}} = (\delta^0_{f(\mathbf{w})} - \frac{1}{2}) \in \{\frac{1}{2}, -\frac{1}{2}\}$, where f is picked from the set of all functions mapping $\{0,1\}^{N_d} \to \{0,1\}$ with uniform distribution. This is equivalent to saying that the 2^{N_d} random variables $\{F_{\mathbf{w}}\}_{\mathbf{w}}$ are independent and identically distributed according to $\Pr\{F_{\mathbf{w}} = \pm\frac{1}{2}\} = \frac{1}{2}$. For ease of notation, let us fix a sequence $(\mathbf{a}_1, \mathbf{x}_1), \dots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ satisfying (B.2.4) and use the short-hand notation $\Gamma^i_{w_i} = \Gamma^i_{w_i}(\mathbf{a}_i, \mathbf{x}_i)$.

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

$$
\begin{aligned}
& \Pr\left\{ \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq \mu \right\} \\
= \; & \Pr\left\{ \nu \left(-\mu + \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right) \geq 0 \right\} \\
= \; & \Pr\left\{ \exp\left(-\nu\mu + \nu \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right) \geq 1 \right\} \\
\leq \; & \mathbb{E}\left[\exp\left(-\nu\mu + \nu \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right) \right] && \text{(B.-20)} \\
= \; & \mathbb{E}\left[\mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \exp\left(\nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right) \right] \\
= \; & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \mathbb{E}\left[\exp\left(\nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right) \right] && \text{(B.-20)} \\
\leq \; & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \mathbb{E}\left[1 + \nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} + \left(\nu F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right)^2 \right] . && \text{(B.-19)}
\end{aligned}
$$

Here $\mathbb{E}$ stands for the average over all $F_{\mathbf{w}}$. In(B.-20) we have used that any positive random variable X satisfies $\Pr\{X \geq 1\} \leq \mathbb{E}[X]$. In(B.-20) we have used that the $\{F_{\mathbf{w}}\}_{\mathbf{w}}$ are independent. Finally, in(B.-19) we have used that

$\mathrm{e}^{\eta} \leq 1 + \eta + \eta^2$, which is only valid if $\eta \leq 1$. Therefore, we must show that

$$\left| \frac{\nu}{2} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \right| \leq 1, \tag{B.-19}$$

which is done below, when setting the value of ν. In what follows we use the chain of inequalities(B.-19), the fact that $\mathbb{E}[F_{\mathbf{w}}] = 0$ and $\mathbb{E}[F^2_{\mathbf{w}}] = 1/4$, bound $1 + \eta \leq \mathrm{e}^{\eta}$ for $\eta \geq 0$, and the definition $\Omega_i^2 = (\Gamma^i_0)^2 + (\Gamma^i_1)^2$:

$$\begin{aligned}
& \Pr\left\{ \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq \mu \right\} \\
\leq\ & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \left(1 + \mathbb{E}[F_{\mathbf{w}}]\, \nu \prod_{i=1}^{N_d} \Gamma^i_{w_i} + \mathbb{E}[F^2_{\mathbf{w}}]\, \nu^2 \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
=\ & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \left(1 + \frac{\nu^2}{4} \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
\leq\ & \mathrm{e}^{-\nu\mu} \prod_{\mathbf{w}} \exp\left(\frac{\nu^2}{4} \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
=\ & \exp\left(-\nu\mu + \sum_{\mathbf{w}} \frac{\nu^2}{4} \prod_{i=1}^{N_d} \left(\Gamma^i_{w_i}\right)^2 \right) \\
=\ & \exp\left(-\nu\mu + \frac{\nu^2}{4} \prod_{i=1}^{N_d} \Omega_i^2 \right)
\end{aligned} \tag{B.-23}$$

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

$$\nu = 2\,\mu \prod_{i=1}^{N_d} \Omega_i^{-2} \ . \tag{B.-23}$$

Note that constraint(B.2.4) implies that the inverse of Ω_i exists. Since we assume $\mu \geq 0$, the initial assumption $\nu \geq 0$ is satisfied by the solution(B.2.4). By substituting(B.2.4) in(B.-23) and rescaling the free parameter μ as

$$\tilde{\mu} = \frac{\mu}{\prod_{i=1}^{N_d} \Omega_i} \ , \tag{B.-23}$$

we obtain

$$\Pr\left\{ \sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq \tilde{\mu} \prod_{i=1}^{N_d} \Omega_i \right\} \leq \mathrm{e}^{-\tilde{\mu}^2} \ , \tag{B.-23}$$

for any $\tilde{\mu} \geq 0$ consistent with condition(B.2.4). We now choose $\tilde{\mu} = 3\sqrt{N_d}$, see Eq.(B.4), getting

$$\Pr\left\{\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq 3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i\right\} \leq \mathrm{e}^{-9N_d} \ . \tag{B.-23}$$

With this assignment, and using (B.2.4) and(B.2.4), condition(B.2.4), yet to be fulfilled, becomes

$$3\sqrt{N_d} \prod_{i=1}^{N_d} \frac{|\Gamma^i_{w_i}|}{\Omega_i} \leq 1 \ , \tag{B.-23}$$

which now holds because of the initial premise(B.4).

Bound(B.2.4) applies to each of the sequences $(\mathbf{a}_1, \mathbf{x}_1), \ldots (\mathbf{a}_{N_d}, \mathbf{x}_{N_d})$ satisfying (B.2.4), and there are at most 4^{5N_d} of them. Hence, the probability that the random function f does not satisfy the bound

$$\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \geq 3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i, \tag{B.-23}$$

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

$$\sum_{\mathbf{w}} F_{\mathbf{w}} \prod_{i=1}^{N_d} \Gamma^i_{w_i} \leq -3\sqrt{N_d} \prod_{i=1}^{N_d} \Omega_i, \tag{B.-23}$$

for at least one sequence satisfying (B.2.4) is also smaller than $1/2$. The lemma now easily follows from these two results. □

B.3. Generality of our results

In this section we make two remarks on the generality of our results. First, we argue that our Theorem still holds if the requirements on the randomness sources are substantially weakened. We conclude that our protocol can amplify essentially any source capable of supplying, as the number of bits it generates grows, an unbounded total amount of randomness. Second, we discuss the meaning of the second step of our protocol. While in principle this step can lead to abortion for some specific ϵ-sources, we argue that this is not relevant in the considered scenarios, neither from a fundamental nor from a cryptographic point of view.

B.3.1. Beyond the ϵ-source

The properties of the ϵ-source appear only in the estimation part of the protocol, that is, in the proof of Lemma B.2. In particular, this lemma exploits bound (B.1.2) on the bit-strings generated by the source. A closer inspection to the proof of this Lemma reveals that it still holds when the distribution $P(x_1, \ldots, x_n|e)$ characterizing the randomness source satisfies

$$G(n,\epsilon) < P(x_1, \ldots, x_n|e) \leq F(n,\epsilon) \ , \tag{B.-23}$$

for any pair of functions $G(n,\epsilon), F(n,\epsilon)$ fulfilling

$$G(n,\epsilon) > 0 \tag{B.-22}$$

$$\lim_{n\to\infty} F(n,\epsilon) = 0 \ . \tag{B.-21}$$

These requirements have an intuitive interpretation. First, the lower bound (B.-22) implies that for any value of e, or any preparation by Eve, all possible combinations of measurement settings have some non-zero probability to appear. Second, condition (B.-21) is necessary to ensure that the seed of randomness is sufficiently large to implement the protocol. In other words, the source of randomness must be able to provide, for an arbitrarily large number of generated bits, an unbounded amount of randomness. Expressed in terms of the min-entropy of the strings, condition (B.-21) implies that $H_{\min}(n,\epsilon) = -\log(F(n,\epsilon))$ can be made arbitrarily large by increasing n. Note that this relaxes considerably the assumptions defining an ϵ-source, which provides exclusively strings with a min-entropy growing linearly with n.

It is an interesting question to explore whether any of these conditions can be relaxed. Our intuition is that growing min-entropy is necessary for randomness amplification and, therefore, we do not expect any significant improvement on the upper bound in (B.3.1). Concerning the lower bound (B.-22), some relaxation is possible. Consider for instance a source that produces two perfect free random bits and then a third one that is the sum of the two. This source violates (B.-22). But it can be trivially mapped into a perfect source by simply discarding the third bit. In general, any source that produces "from time to time" ϵ-random bits is good for our purposes, as it can be mapped into a source satisfying the previous requirements by computing parities of large enough bit-strings. Sources that do not satisfy this condition are such that for all i there exists a value of Eve's variable e and the other symbols, x_j with $j \neq i$, such that $P(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, x_n, e) = 1$. Whether randomness amplification is possible out of these sources, and even whether these sources can be interpreted as proper random sources, are interesting questions that deserve further investigation.

B.3.2. Step 2 of the protocol

Step 2 in our protocol demands that the fraction of events in which the settings needed for the 5-party Mermin inequality, denoted by $\mathcal{X}$, are generated by the ϵ-source is larger than $N/3$. First of all, it is worth mentioning that the choice of the factor $1/3$ is arbitrary and, in fact, any linear function of N is valid for our protocol. Let us group the bits provided by the source in N groups of five bits $\mathbf{x}_i = (x_{5i+1}, \ldots, x_{5i+5})$. Recall from chapter 4 that the set $\mathcal{X}$ defines the inputs appearing in the five-party Bell inequality. Let us note as $N_{\mathcal{X}}(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ the number of elements in the list $\mathbf{x}_1, \ldots, \mathbf{x}_N$ belonging to $\mathcal{X}$. In order to have a large probability of not aborting the protocol, the source must satisfy

$$P\left(N_{\mathcal{X}}(\mathbf{x}_1, \ldots, \mathbf{x}_N) > \frac{N}{3}\right) \approx 1 \ , \tag{B.-21}$$

for sufficiently large N. This ensures that the step 2 of the protocol (see chapter 4) passes with high probability. Note that there are ϵ-sources that pass this step of the protocol with exponentially small probability, for instance when they are always biased towards the same bit value. But, as we discuss in what follows, the requirement of step 2 is very natural.

First, from a fundamental point of view, the protocol for randomness amplification only makes sense for sources that look perfectly random. In fact, it would not be very meaningful to pose the question of full randomness amplification if all the randomness one could see is biased towards a given value. Now, if the source looks random, step 2 will be automatically satisfied with very high probability. Formally, a source looks random if $P(x_1, \ldots, x_n) = \frac{1}{2^n}$. Taking into account that half of the possible values that $\mathbf{x}_i$ can take belong to the set $\mathcal{X}$, typical sequences of such source fulfill $\frac{N}{2} + c\sqrt{N} \leq N_{\mathcal{X}}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \leq \frac{N}{2} + c\sqrt{N}$, for some constant c. In (B.3.2) we allow for a linear deviation from the typical sequences by requiring only that $\frac{N}{3}$ of the strings belong to $\mathcal{X}$. Therefore, with arbitrarily high probability, the source fulfills (B.3.2).

Second, from a cryptographic point of view, abortion is not a particularly relevant issue, as it is always possible for an eavesdropper to make the protocol abort with high probability, for instance by preparing a deterministic source. The protocol is meaningful from a cryptographic point of view when it guarantees that the probability that an ϵ- source passes the protocol and does not get amplified is exponentially small, as it is the case for the present protocol. Finally, note that imposing step 2 does not imply any loss on the predictability power of a non-signalling eavesdropper preparing the ϵ-source. In fact, consider a source that does not pass step 2. The eavesdropper can also prepare a second source almost identical to the previous but where the roles of zeros and ones are

exchanged. Now the eavesdropper correlates these two sources with an extra bit e_p in her possession that has the same probability of being equal to zero or one. Clearly, the resulting ϵ-source (i) gives the same predictability power to the eavesdropper, who has just to measure the bit e_p to know the actual prepared source and (ii) passes step 2 of the protocol.

B.4. Final remarks

The main goal was to prove full randomness amplification. In this Appendix, we have shown how our protocol, based on quantum non-local correlations, achieves this task. Unfortunately, we are not able to provide an explicit description of the function $f : \{0,1\}^{N_d} \to \{0,1\}$ which maps the outcomes of the black boxes to the final random bit k; we merely show its existence. Such function may be obtained through an algorithm that searches over the set of all functions until it finds one satisfying (B.4). The problem with this method is that the set of all functions has size 2^{N_d}, which makes the search computationally costly. However, this problem can be fixed by noticing that the random choice of f in the proof of Lemma B.4 can be restricted to a four-universal family of functions, with size polynomial in N_d. This observation will be developed in future work.

A more direct approach could consist of studying how the randomness in the measurement outcomes for correlations maximally violating the Mermin inequality increases with the number of parties. We solved linear optimization problems similar to those used in lemma 4.1 in chapter 4 , which showed that for 7 parties Eve's predictability is $2/3$ for the permutationally symmetric function of 5 bits defined by $f(00000) = 0$, $f(01111) = 0$, $f(00111) = 0$ and $f(\mathbf{x}) = 1$ otherwise. Note that this value is lower than the earlier $3/4$ and also that the function is different from the majority-vote. We were however unable to generalize these results for an arbitrary number of parties, which forced us to adopt a less direct approach. Note in fact that our protocol can be interpreted as a huge multipartite Bell test from which a random bit is extracted by classical processing of some of the measurement outcomes.

We conclude by commenting on the reason why randomness amplification becomes possible using non-locality when it was an impossible task classically [SV86]. In our protocol, we use the bits produced by the SV-source to run a Bell test on a particular causal structure. The observed Bell inequality violation certifies some randomness in the measurement outcomes. However the observed Bell violation also impose a non-trivial structure on the correlations among quintuplets, as non-local correlations have to satisfy some monogamy constraints to be compatible with the no-signalling principle, which permits

us to certify that the distilled bit is fully uncorrelated. There already exist several protocols, both in classical and quantum information theory, in which imperfect randomness is processed to generate perfect (or arbitrarily close to perfect) randomness. However, all these protocols, e.g. two-universal hashing or randomness extractors, always require additional good-quality randomness to perform such distillation. On the contrary, if the initial imperfect randomness has been certified by a Bell inequality violation, the distillation procedure can be done with a deterministic hash function (see [Mas09] or Lemma B.1 above). This property makes Bell-certified randomness fundamentally different from any other form of randomness, and is the key for the success of our protocol.

C. Proofs of certification of fully intrinsic randomness

C.1. Proof of lemma 5.1

Here we prove the lemma upon which its based theorem 5.2 which constitutes the main result of chapter 5. It is basically a generalization of the the proof for $N = 3$ given in the said chapter. We would like to prove that the function f defined in chapter 5, satisfies the property:

$$P(f(\mathbf{a}) = h_N|\mathbf{x_m}) \geq 1/2 \tag{C.0}$$

for any N-partite distribution (odd N) that maximally violates the Mermin inequality. As in the tripartite case, in order to prove the result we (I) express condition (C.1) in terms of some correlators and (II) use positivity conditions from the swapped input to prove the inequality.

An N-partite no-signalling probability distribution $P(\mathbf{a}|\mathbf{x})$ with inputs $\mathbf{x} \in \{0,1\}^N$ and outputs $\mathbf{a} \in \{+1,-1\}^N$ can be parameterized in terms of correlators as,

$$P(\mathbf{a}|\mathbf{x}) = \frac{1}{2^N}\Big(1 + \sum_{i=1}^{N} a_i\langle x_i\rangle + \sum_{i<j} a_i a_j\langle x_i x_j\rangle + \sum_{i<j<k} a_i a_j a_k\langle x_i x_j x_k\rangle + \ldots + a_1 a_2 \ldots a_N\langle x_1 x_2 \ldots x_N\rangle\Big). \tag{C.1}$$

Restricting $P(\mathbf{a}|\mathbf{x})$ to those maximally violating the N-partite Mermin inequality is equivalent to requiring all correlators of input strings of odd parity to take their extremal values. Namely, we have,

$$\langle x_1 x_2 \ldots x_N\rangle = (-1)^{(-1+\sum_{i=1}^{N} x_i)/2}, \tag{C.1}$$

for all N-point correlators satisfying $\sum_{i=0}^{N} x_i = 1 \bmod 2$. For example, the correlator $\langle 0,0,\ldots,1\rangle = 1$, and similarly for all permutations. Also, $\langle 0,0,\ldots,0,1,1,1\rangle = -1$ as well as for for all permutations, etc. In the following we will use the notation $\langle . \rangle_k$ to denote a k-point correlator. The input combination used to extract

randomness is a generalization of the tripartite case and denoted by $\mathbf{x_m} = (0, 0, \ldots, 0, 1)$. The corresponding N-point correlator satisfies $\langle 0, 0, \ldots, 0, 1\rangle = 1$ for all N. The latter implies two useful relations:

1. Half the total outcomes vanish. In particular these are the terms for which the product of outcomes is -1 *i.e.* $P(\prod_{i=1}^{N} a_i = -1|\mathbf{x_m}) = 0$.

2. $\langle . \rangle_{N-k} = \langle . \rangle_k$ for all $1 \leq k \leq (N-1)/2$ where the correlators $\langle . \rangle_{N-k}$ and $\langle . \rangle_k$ are complementary in the input $\mathbf{x_m}$.

One can use these in Eqn. C.1 to express $P(\mathbf{a}|\mathbf{x_m})$ in terms of only the first $(N-1)/2$-point correlators as,

$$P(\mathbf{a}|\mathbf{x_m}) = \frac{1}{2^{N-1}}(1 + \sum a_i\langle x_i\rangle + \sum a_i a_j \langle x_i x_j\rangle + \cdots + \sum a_i a_j \cdots a_p \langle x_i x_j \cdots x_p\rangle_{(N-1)/2}) \tag{C.2}$$

where $a_1 \cdot a_2 \cdot a_3 \ldots a_N = +1$ since $P(\mathbf{a}|\mathbf{x_m}) = 0$ when $a_1 \cdot a_2 \cdot a_3 \ldots a_N = -1$.

C.1.1. Expressing the target inequality in terms of correlators

As mentioned, our first goal is to express Eq.(C.1) as a function of some correlators. Let us recall the function we use in our main theorem,

$$f(\mathbf{a}) = \begin{cases} +1 & n_-(\mathbf{a}) = (4j+2); \quad \text{with } j \in \{0, 1, 2, \ldots\} \\ -1 & \text{otherwise} \end{cases} \tag{C.1}$$

where $n_-(\mathbf{a})$ denotes the number of results in $\mathbf{a}$ that are equal to -1.

It turns out that the quantity (Eq. (C.1)) we would like to calculate, namely, $P(f(\mathbf{a}) = h_N|\mathbf{x}_m) - 1/2$ can be equivalently expressed as $h_N \cdot (P(f(\mathbf{a}) = +1|\mathbf{x}_m) - 1/2)$. The latter form is convenient since the function only takes value $+1$ for all N.

We proceed to express the latter in terms of correlators (as in the proof for three parties in Chapter 5),

$$(h_N \cdot P(f(\mathbf{a}) = +1|\mathbf{x}_m) - 1/2) = 2^{-(N-1)}\boldsymbol{\alpha}' \cdot \mathbf{c}, \tag{C.1}$$

where

$$\begin{aligned} \boldsymbol{\alpha}' &= h_N \cdot (\alpha_0 - 2^{N-2}, \alpha_1, \alpha_2, \ldots, \alpha_{(N-1)/2}) \\ \mathbf{c} &= \left(1, \sum_{\mathcal{S}^1}\langle . \rangle_1, \sum_{\mathcal{S}^2}\langle . \rangle_2, \ldots, \sum_{\mathcal{S}^{(N-1)/2}}\langle . \rangle_{(N-1)/2}\right) \end{aligned} \tag{C.1}$$

Note that, since the function f symmetric under permutations, the vector $\mathbf{c}$ consists of the different sums of all k-point correlators, denoted by $\mathcal{S}^k$, where k ranges from 0 to $(N-1)/2$ because of Eq. (C.2). The vector $\boldsymbol{\alpha}'$ is the vector of coefficients for each sum of correlators. Our next goal is to compute this vector.

Recall that function f is such that $f(\mathbf{a}) = +1$ if $n_-(\mathbf{a}) = 4j+2$ for any $j \in \mathbb{N} \cup \{0\}$. By inspection, the explicit values of α_i can be written as

$$\alpha_i = \sum_{r=0}^{i} (-1)^r \binom{i}{r} \sum_{j\geq 0} \binom{n-i}{4j+2-r}. \tag{C.1}$$

For example, $\alpha_0 = \sum_{j\geq 0} \binom{n}{4j+2}$ as one would expect since α_0 simply counts the total number of terms $P(\mathbf{a}|\mathbf{x}_m)$ being summed to obtain $P(f(\mathbf{a}) = +1|\mathbf{x}_m)$.

Making use of the closed formula $\sum_{j\geq 0} \binom{n}{rj+a} = \frac{1}{r} \sum_{k=0}^{r-1} \omega^{-ka}(1+\omega^k)^n$ [BCK10], where $\omega = e^{i2\pi/r}$ is the r^{th} root of unity, we can simplify the second sum appearing in Eq. C.1.1. Finally we recall that the phase h_N was defined (in Chapter 5) to be $h_N = \sqrt{2}\cos{(N+4)\pi/4}$. Putting all this together and performing the first sum in Eq. (C.1.1) gives us,

$$\alpha'_i = 2^{\frac{N-3}{2}} \left(-2 \cos \frac{(N-2i)\pi}{4} \cos \frac{(N+4)\pi}{4} \right) \tag{C.1}$$

Notice that the term in the parenthesis is a phase taking values in the set $\{+1, -1\}$ since N is odd while the amplitude is independent of N. Thus, we can simplify Eqn. (C.1.1) for even and odd values of i as,

$$\alpha'_i = \begin{cases} 2^{(N-3)/2}(-1)^{\frac{N-i}{2}} & i \text{ odd} \\ \\ 2^{(N-3)/2}(-1)^{\frac{i}{2}} & i \text{ even} \end{cases} \tag{C.1}$$

Thus, to prove that f possesses the property $h_N \cdot (P(f(\mathbf{a}) = +1|\mathbf{x_m}) - 1/2) \geq 0$ necessary to proving the main theorem is equivalent to proving

$$\boldsymbol{\alpha}' \cdot \mathbf{c} \geq 0, \tag{C.1}$$

for $\mathbf{c}$ as defined in Eqn. (C.1.1) and for the values of $\boldsymbol{\alpha}'$ given by Eqn. (C.1.1). This is the task of the following section, where we show that it follows from positivity constraints on $P(\mathbf{a}|\mathbf{x})$.

C.1.2. Proving the target inequality from positivity constraints

We show that positivity conditions derived from the swapped input $\overline{\mathbf{x}}_\mathbf{m} = (1, 1, \ldots, 1, 0)$ may be used to show $\boldsymbol{\alpha}' \cdot \mathbf{c} \geq 0$. Notice that the components of

$\overline{\mathbf{x}}_\mathbf{m}$ and $\mathbf{x}_\mathbf{m}$ are opposite, ie. $\{\overline{\mathbf{x}}_\mathbf{m}\}_i = \{\mathbf{x}_\mathbf{m}\}_i \oplus 1$ for all i. In the following we will repeatedly use the Mermin conditions of Eqn. (C.1).

We start by summing the positivity conditions $P(+++\cdots+-|\overline{\mathbf{x}}_\mathbf{m}) \geq 0$ and $P(---\cdots-+|\overline{\mathbf{x}}_\mathbf{m}) \geq 0$. Using Eqn. (C.1), one can easily see that upon summing, all k-point correlators for *odd* k are cancelled out since these are multiplied by coefficients (products of a_is) that appear with opposite signs in the two positivity expressions. In contrast, k-point correlators for *even* k add up since they are multiplied by coefficients that appear with the same sign in the two expressions. For example, N being odd, the full correlator always cancels out while the $(N-1)$-point correlators always appear.

This leaves us with an expression containing only the even-body correlators,

$$1+\sum_{i<j} a_i a_j \langle x_i x_j\rangle + \sum_{i<j<k<l} a_i a_j a_k a_l \langle x_i x_j x_k x_l\rangle + \cdots + \sum a_i \ldots a_p \underbrace{\langle x_i \ldots x_p\rangle}_{(N-1)\text{-pt. corr}} \geq 0. \tag{C.1}$$

Note once again, that this inequality is derived from the so-called swapped input $\overline{\mathbf{x}}_\mathbf{m}$. We aim to cast it in a form that can be compared directly with Eqn. (C.1.1), which comes from the chosen Mermin input $\mathbf{x}_\mathbf{m}$. To this end, we need to convert Eqn. (C.1.2) to an expression of the form,

$$(\beta_0, \beta_1, \ldots \beta_{(N-1)/2}).\left(1, \sum\langle.\rangle_1, \ldots, \sum\langle.\rangle_{(N-1)/2}\right) \geq 0 \tag{C.1}$$

We first highlight the similarities and differences between the two preceding expressions, namely, the one we have *i.e.* Eqn. (C.1.2) and the one we want, *i.e.* Eqn. (C.1.2). Each contains $(N-1)/2$ distinct classes of terms. However the former contains only even k-point correlators for $k = 2$ to $(N-1)$ while the latter contains all terms from $k = 1$ to $(N-1)/2$. Thus, terms of Eq. (C.1.2) must be mapped to ones in Eqn. (C.1.2). Moreover, since the point of making this mapping is to finally compare with Eqn. (C.1.1), we also note that the correlators appearing in Eqn. (C.1.2) are locally swapped relative to those appearing in Eqn. (C.1.1). Thus, our mapping must also convert correlators of the swapped input into those corresponding to the chosen input.

We demonstrate next that one may indeed transform the inequality (C.1.2) into the inequality (C.1.2) satisfying both the demands above. To this end, all the *even* k-point correlators (for $k \geq \frac{N-1}{2}$) appearing in Eqn. (C.1.2) are mapped to odd $(N-k)$-point correlators in Eqn. (C.1.2). Likewise, all the even k-point correlators (for $k < \frac{N-1}{2}$) of the swapped input appearing in Eqn. (C.1.2) are mapped to the corresponding k-point correlators of the chosen input in Eqn. (C.1.2).

These mappings make systematic use of the Mermin conditions Eqn. (C.1) and are made explicit in the following section.

Even-point correlators

Consider a $2k$-point correlator where $2k \leq (N-1)/2$. The correlators are of two forms and we show how they are transformed in each case:

- $\langle 11\ldots1\rangle_{2k}$. We would like to map this to the correlator $\langle 00\ldots0\rangle_{2k}$ appearing in $\mathbf{x_m}$. We achieve the mapping by completing each to the corresponding Mermin full-correlators $\langle \underbrace{11\ldots1}_{2k}\underbrace{100\ldots0}_{(N-2k)}\rangle_N = (-1)^k$ and $\langle \underbrace{00\ldots0}_{2k}\underbrace{100\ldots0}_{(N-2k)}\rangle_N = (-1)^0 = 1$. From the signs, we have the relation, $\langle 11\ldots1\rangle_{2k} = (-1)^k\langle 00\ldots0\rangle_{2k}$

- $\langle 11\ldots10\rangle_{2k}$, which we would like to map to $\langle 00\ldots01\rangle_{2k}$. Using the same ideas we get $\langle \underbrace{11\ldots10}_{2k}\underbrace{110\ldots0}_{(N-2k)}\rangle_N = (-1)^k$ and $\langle \underbrace{00\ldots01}_{2k}\underbrace{110\ldots0}_{(N-2k)}\rangle_N = (-1)^1 = -1$. Thus, giving us the relation $\langle 11\ldots10\rangle_{2k} = (-1)^{k+1}\langle 00\ldots01\rangle_{2k}$.

By inspection one can write the relationship

$$\underbrace{a_1a_2\ldots a_{2k}}_{\text{even}}\underbrace{\langle x_1x_2\ldots x_{2k}\rangle}_{\text{cor in }\overline{\mathbf{x}}_\mathbf{m}} = (-1)^k \underbrace{\langle x_1x_2\ldots x_{2k}\rangle}_{\text{cor in }\mathbf{x_m}(desired)}$$

for correlators of either form discussed above on multiplying with their corresponding coefficients. Since we have finally converted to the desired correlators of the chosen input $\overline{\mathbf{x}}$, we can read off β_i as the corresponding phase. Thus, $\beta_i = (-1)^{i/2}$ for even i.

Odd-point correlators

Consider now a $2k$-point correlator where $2k \geq (N-1)/2$. The correlators are again of two forms and may be transformed to the required $(N-2k)$-point correlators in each case. The only difference from before is that the two correlators are now complementary to each other in the swapped input.

Since the details are similar, we simply state the final result $\beta_i = (-1)^{(N-i)/2}$ for odd i.

The final expression thus reads,

$$\beta_i = \begin{cases} (-1)^{\frac{N-i}{2}} & i \text{ odd} \\ \\ (-1)^{\frac{i}{2}} & i \text{ even} \end{cases} \tag{C.1}$$

Thus, the values of β given in Eqs. (C.1.2) exactly match the ones for α_i' (up to the constant factor) given in Eqn. C.1.1. Together with the correlators matching those in $\mathbf{c}$, it proves that f satisfies the required $\boldsymbol{\alpha}' \cdot \mathbf{c} \geq 0$ and hence the full result.

As a last remark, note that, the input combination appearing in the inequality that we choose is arbitrary. Although everything is proved for the particular measurement combination $\mathbf{x_m}$, it is easy to see that, because of the symmetries of the Mermin inequalities, Lemma C.1 holds with a completely analogous proof for all inputs appearing in the inequality whose correlator is equal to $+1$ according to the Mermin conditions for maximal violation of Eqn. (C.1). If the combination chosen has an associated correlator equal to -1 according to Eqn. (C.1), the lemma holds for the related function $f'(\mathbf{a})$:

$$f'(\mathbf{a}) = \begin{cases} +1 & n_+(\mathbf{a}) = (4j+2); \quad \text{with } j \in \{0,1,2,\ldots\} \\ \\ -1 & \text{otherwise} \end{cases} \tag{C.1}$$

where $n_+(\mathbf{a})$ denotes the number of results in $\mathbf{a}$ that are equal to $+1$. Note that this function is equivalent to, first locally swap all the outputs, and then apply the original function $f(\mathbf{a})$. The proof follows since after locally swapping all the outputs of all parties (odd number), the correlator associated with this measurement combination changes sign.

C.2. Proof that all distributions in decomposition maximally violate the Mermin inequality

We end by proving the claim made during the proof of theorem 5.2 in chapter 5, section 5.2.2 that if an observed probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ violates maximally and algebraically the corresponding Mermin inequality, all the no-signaling components $P_e^{\text{ex}}(\mathbf{a}|\mathbf{x})$ present in its preparation must also algebraically violate the inequality.

We recall that the decomposition appears in the definition of intrinsic randomness given by,

$$G_{\text{int}}(g, \mathbf{x}, P_{\text{obs}}) = \max_{\{p(e|\mathbf{x}), P_e^{\text{ex}}\}} \sum_e p(e|\mathbf{x}) G_{\text{obs}}(g, \mathbf{x}, P_e^{\text{ex}})$$

subject to:

$$\sum_e p(e|\mathbf{x}) P_e^{\text{ex}}(\mathbf{a}|\mathbf{x}) = P_{\text{obs}}(\mathbf{a}|\mathbf{x})$$

$$p(\mathbf{x}|e) \geq \delta \;\text{ with } \delta > 0 \;\; \forall\, \mathbf{x}, e$$

Since P_{obs} algebraically violates the Mermin inequality, this definition imposes stringent conditions on the correlators of P_{obs} satisfying the Mermin condition (C.1), namely that,

$$\langle x_1 \dots x_N \rangle_{P_{\text{obs}}} = \pm 1 = \sum_e p(e|x_1, \dots, x_N) \, \langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}} \tag{C.-1}$$

where by normalization $\sum_e p(e|x_1, \dots, x_N) = +1$ and $-1 \leq \langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}} \leq +1$. Note that condition $p(\mathbf{x}|e) \geq \delta$ for all $\mathbf{x}, e$ for $\delta > 0$ can be inverted using the Bayes' rule to obtain $p(e|\mathbf{x}) > 0$ for all $\mathbf{x}, e$. Now is clear by convexity that the condition $p(\mathbf{x}|e) \geq \delta$ (denying *absolute* relaxation of freedom of choice) implies that all the correlator $\langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}}$ appearing in the Mermin inequality must also necessarily satisfy $\langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}} = \pm 1$ for all e thus maximally violating the Mermin inequality. In fact it is also clear that this constraint on $p(\mathbf{x}|e)$ is strictly necessary to ensure that the decomposition correlations satisfy maximal Mermin violation. To see this, suppose $p(\mathbf{x}|e_0) = 0$, then the corresponding $\langle x_1 \dots x_N \rangle_{P_{e_0}^{\text{ex}}}$ is fully unconstrained while satisfying Eq. (C.2).

D. Proofs of results of chapter 6

D.1. Proof of Result 2

In this section we show that it is possible to obtain N bits of global randomness for all N. In Ref. [DPA13] it was shown how to achieve and certify N bits of global randomness for all odd N. Here, we show that there is a Bell inequality in the $(N, 2, 2)$ setting for all N (which is a generalization of the Mermin inequality first studied in [HCLB11]) which if maximally violated, gives N bits of global randomness. We use the tools developed in Ref. [DPA13] to obtain maximal randomness based on the symmetries of the inequality we use.

First, we need to introduce some notation. As standard in the literature, we introduce the n-party correlators where $n \leq N$. Take a subset $J_n \subseteq \{1, 2, ..., N\}$ of n parties from all N parties. Then associated with this subset and a string of inputs $x_{J_n} = \{x_j | j \in J_n\}$, a string of outputs $a_{J_n} = \{a_j | j \in J_n\}$ and a marginal probability distribution $P(a_{J_n}|x_{J_n})$. We then define the correlators to be

$$\langle x_{J_n} \rangle := 2 \left(\sum_{a_{J_n}} \alpha P(a_{J_n}|x_{J_n}) \right) - 1, \tag{D.0}$$

with $\alpha = 1 + \sum_{k \in J_n} a_k \bmod 2$. We can define the full joint probabilities in terms of these correlators as

$$P(\mathbf{a}|\mathbf{x}) = \frac{1}{2^N} \sum_{J_n} (-1)^{\sum_{k \in J_n} a_k} \langle x_{J_n} \rangle, \tag{D.0}$$

where we take a sum over all 2^N subsets of N parties (including the empty set).

The Bell inequality that will concern is the following inequality discussed in Ref. [HCLB11]:

$$\sum_{\mathbf{x}} (-1)^{f(\mathbf{x})} \delta_0^{g(\mathbf{x})} \langle x_{J_n} \rangle \leq \epsilon < 2^{N-1}, \tag{D.0}$$

where $f(\mathbf{x}) = \sum_{j=1}^{N-1} x_j \left(\sum_{k=j+1}^{N} x_k \right) \bmod 2$ and $g(\mathbf{x}) = \sum_{j=1}^{N} x_j \bmod 2$. As indicated the upper-bound for local hidden variables ϵ is strictly less than 2^{N-1}, the number of terms in the sum. Crucially, quantum mechanics can violate

this inequality and achieve the algebraic upper bound of 2^{N-1} as shown in Ref. [HCLB11] using a Greenberger-Horne-Zeilinger state. Also, there is only one probability distribution that maximally violates this inequality as can be shown by applying the techniques of Ref. [FFW11] or by a self-testing argument due to [MS13]. In Ref. [DPA13] this property of *uniqueness* of a probability distribution maximally violating an inequality was used to prove that global randomness can be generated from a Bell test.

We need to use an input $\mathbf{x}'$ that does not appear in the left-hand-side of (D.1), since the outputs of measurements for inputs in (D.1) will be highly correlated, and thus not random. We need to show that for input $\mathbf{x}'$, the probability $P(\mathbf{a}|\mathbf{x}')$ in (D.1) is equal to $\frac{1}{2^N}$ for all $\mathbf{a}$. This occurs if all correlators satisfy $\langle x_{J_n}\rangle = 0$ for all (non-empty) subsets J_n and $\langle x_{J_n}\rangle = 1$ for $J_n = \emptyset$, the empty set when $n = 0$. The aim of this section is to show that this is true.

To do this, we utilize the tools in Ref. [DPA13] where we perform transformations on the data obtained in a Bell test that do not affect the correlators that appear in the Bell inequality of (D.1). These transformations affect correlators that do appear in the inequality. If we take the unique probability distribution that maximally violates (D.1) then, under these transformations, it still violates the same inequality maximally. If we call the original probability distribution P with elements $P(\mathbf{a}|\mathbf{x})$ and the transformed distribution P', then $P = P'$, and so all correlators resulting from these two distributions must be equal as well. If one of these symmetry transformations is to flip an outcome of a measurement depending on the choice of input then this can alter correlators, e.g. $a_1 \to a_1 \oplus x_1$, then all correlators $\langle x_{J_n}\rangle$ that contain $x_1 = 1$ have their sign flipped as $\alpha = 1 + \sum_{k\in J_n} a_k \bmod 2 \to 2 + \sum_{k\in J_n} a_k \bmod 2$. However, due to uniqueness of quantum violation this implies that the correlators before and after the transformation are equal, so in the case that the transformation flips the sign of the correlator then $\langle x_{J_n}\rangle = -\langle x_{J_n}\rangle = 0$. This thus demonstrates a way to show that correlators are zero for particular distributions.

For clarity we introduce the notation to show when a correlator's sign is flipped. Our symmetry operations are captured by an n-length bit-string $\mathbf{s}$, where if the jth element s_j is zero, then we flip a_j for the choice of input $x_j = 0$, and if $s_j = 1$, then we flip a_j for choice of input $x_j = 1$. Then the correlator $\langle x_{J_n}\rangle$ under the symmetry transformation described by s is mapped to $(-1)^{N-H(\mathbf{x},\mathbf{s})}\langle x_{J_n}\rangle$ where $H(\mathbf{x},\mathbf{s})$ is the Hamming distance between the bit-strings $\mathbf{s}$ and $\mathbf{x}$: the number of times $x_j \neq s_j$ for bit-strings $\mathbf{x}$, $\mathbf{s}$. Another way of writing the Hamming distance is

$$H(\mathbf{x},\mathbf{s}) = \sum_{j=1}^{N} x_j + s_j \bmod 2 = \sum_{j=1}^{N} x_j + s_j - 2s_j x_j. \tag{D.0}$$

We now return to the Bell inequality in (D.1) and focus on even N. We apply N transformations described by the bit-strings $\mathbf{s}$: $(0,0,...,0)$ (the all-zeroes bit-string) and the $(N-1)$ bit-strings $\mathbf{s}$ that all have $s_N = 1$, and only one other element being equal to one, e.g. $(1,0,0,...,1)$ or $(0,1,0,...,1)$. All of these bit-strings have an even number of ones, therefore $\sum_{j=1}^{N} s_j = 2k$ for $k \in \{0,1\}$. For correlators in the inequality of (D.1), the inputs $\mathbf{x}$ satisfy $\sum_{j=1}^{N} x_j \bmod 2 = 0$, so that $\sum_{j=1}^{N} x_j = 2k'$ for k' being some integer. Therefore all the correlators $\langle x \rangle$ that appear in (D.1) are mapped to $(-1)^{2(k+k'-\sum_{j=1}^{N} s_j x_j)}\langle x \rangle = \langle x \rangle$ and thus the transformation does not alter the inequality.

To obtain N bits of global randomness for even N, we choose the input $\mathbf{x}' = (1,1,...,1,0)$, the bit-string of all-ones except $x'_N = 0$. This input does not appear in (D.1) and indeed $\sum_{j=1}^{N} x_j = 2k'+1$ for some integer k', therefore the above transformations map $\langle x' \rangle$ to $(-1)^{1+2(k+k'-\sum_{j=1}^{N} s_j x_j)}\langle x' \rangle = -\langle x' \rangle$. Due to the uniqueness of the probability distribution maximally violating the Bell inequality, $\langle x' \rangle = -\langle x' \rangle = 0$.

We now need to show that all correlators $\langle x'_{J_n} \rangle$ where x'_{J_n} is the string of $n < N$ elements from $\mathbf{x}' = (1,1,...,1,0)$ for a sub-set J_n are zero. We consider the Hamming distance $H(x'_{J_n}, s_{J_n})$ between x'_{J_n} and the corresponding string s_{J_n} of elements of $\mathbf{s}$ where s_j is in s_{J_n} if $j \in J_n$. Immediately we see that for at least one string s_{J_n}, the Hamming distance is $H(x'_{J_n}, s_{J_n}) = (n-1)$. Therefore, there is at least one transformation $\mathbf{s}$ that maps $\langle x'_{J_n} \rangle$ to $(-1)^{n-(n-1)}\langle x'_{J_n} \rangle = -\langle x'_{J_n} \rangle$ for all J_n. Again, given that all correlators should be equal after the transformation we have that $\langle x'_{J_n} \rangle = -\langle x'_{J_n} \rangle = 0$.

To summarize, we have shown that all correlators that appear in (D.1) for the input $\mathbf{x}' = (1,1,...,1,0)$ are equal to zero if the probability distribution that produces them maximally violates the inequality in (D.1). This therefore implies that $P(\mathbf{a}|\mathbf{x}') = \frac{1}{2^N}$ for all $\mathbf{a}$ and for all even N. For this input $\mathbf{x}'$ we obtain N bits of global randomness. We can use another inequality to obtain N bits of global randomness for odd N as shown in Ref. [DPA13]. Therefore, we can obtain N bits of global randomness for all N. We have used the fact that there is a unique *quantum* violation of the inequality in (D.1). However, for more general theories this may not be the case.

D.2. Proof of Result 3

We present a proof that it is possible to certify 3 bits of global randomness for a set of correlations that is strictly larger than the quantum set Q. We call this set Q^{1+ABC} in the terminology of the multipartite generalization of

the Navascués-Pironio-Acín hierarchy of correlations that can be characterized through semi-definite programming [NPA07]. We prove this result utilising the tripartite Mermin inequality [Mer90], so we are therefore in the $(3, 2, 2)$ scenario.

We first recall from Ref. [NPA07] that correlations $P(a_1, a_1, a_3|x_1, x_1, x_3)$ are contained in the set Q^{1+ABC} if there exists a pure quantum state $|\psi\rangle$, and projectors $\{E^{a_1}_{x_1}, F^{a_1}_{x_1}, G^{a_3}_{x_3}\}$ labelled by inputs $x_j \in \{0, 1\}$ and outputs $a_j \in \{0, 1\}$, such that

1. *(Hermiticity)* – $(E^{a_1}_{x_1})^\dagger = E^{a_1}_{x_1}$, $(F^{a_1}_{x_1})^\dagger = F^{a_1}_{x_1}$, and $(G^{a_3}_{x_3})^\dagger = G^{a_3}_{x_3}$ for all x_j and a_j

2. *(Normalization)* – $\sum_{a_1} E^{a_1}_{x_1} = \mathbb{I}$, $\sum_{a_1} F^{a_1}_{x_1} = \mathbb{I}$, and $\sum_{a_3} G^{a_3}_{x_3} = \mathbb{I}$ for all x_j

3. *(Orthogonality)* – $E^{a_1}_{x_1} E^{a'_1}_{x_1} = \delta^{a_1}_{a'_1} E^{a_1}_{x_1}$, $F^{a_1}_{x_1} F^{a'_1}_{x_1} = \delta^{a_1}_{a'_1} F^{a_1}_{x_1}$, and $G^{a_3}_{x_3} G^{a'_3}_{x_3} = \delta^{a_3}_{a'_3} G^{a_3}_{x_3}$ for all x_j,

such that probabilities are $P(a_1, a_1, a_3|x_1, x_1, x_3) = \langle\psi|E^{a_1}_{x_1} F^{a_1}_{x_1} G^{a_3}_{x_3}|\psi\rangle$. In addition to these general constraints, linear combinations of these probabilities are elements of a positive semidefinite matrix $\Gamma_{1+ABC} \succeq 0$. We choose a specific positive semidefinite matrix with elements $[\Gamma_{1+ABC}]_{ij} = \langle\psi|\mathcal{O}^\dagger_i \mathcal{O}_j|\psi\rangle$ where $\mathcal{O}_l \in \{\mathbb{I}, \{A_i\}, \{B_j\}, \{C_k\}, \{A_i B_j C_k\}\}$ for $A_i = E^0_i - E^1_i$, $B_j = F^0_j - F^1_j$ and $C_k = G^0_k - G^1_k$. Therefore the matrix Γ_{1+ABC} is a 15-by-15 matrix with each $\mathcal{O}_l$ labelling a row or column. We can now make several observations: $\mathcal{O}_l \mathcal{O}_l = \mathbb{I}$ for all $\mathcal{O}_l$ therefore $(\mathcal{O}_l)^\dagger = \mathcal{O}_l$; $\langle x_1 x_1 x_3\rangle = \langle\psi|A_{x_1} B_{x_1} C_{x_3}|\psi\rangle$; $\langle x_1 x_1\rangle = \langle\psi|A_{x_1} B_{x_1}|\psi\rangle$; $\langle x_1 x_3\rangle = \langle\psi|A_{x_1} C_{x_3}|\psi\rangle$; $\langle x_1 x_3\rangle = \langle\psi|B_{x_1} C_{x_3}|\psi\rangle$; $\langle x_1\rangle = \langle\psi|A_{x_1}|\psi\rangle$; $\langle x_1\rangle = \langle\psi|B_{x_1}|\psi\rangle$; and $\langle x_3\rangle = \langle\psi|A_{x_3}|\psi\rangle$. Here we utilized the notation introduced in the previous section D.1. Finally, the set of quantum correlations Q is a subset of Q^{1+ABC} since the former can be recovered from the latter by imposing more constraints on the projectors. It can also be shown that Q is a strict subset of Q^{1+ABC} for all possible scenarios (N, M, d).

Now that we have defined the set Q^{1+ABC} of correlations that concerns us, we return to the issue of randomness certification. We wish to show that for correlations in this set that maximally violate the tripartite Mermin inequality [Mer90]

$$\langle 001\rangle + \langle 010\rangle + \langle 100\rangle - \langle 111\rangle \leq 2, \tag{D.0}$$

$P(\mathbf{a}|\mathbf{x}_0) = \frac{1}{8}$ for all $\mathbf{a}$ for a particular input $\mathbf{x}_0$. We choose this input to be $\mathbf{x}_0 = (0, 0, 0)$ but it will turn out that we could choose any input $\mathbf{x}$ that does not appear in the Mermin inequality. The maximal violation of the Mermin inequality is 4 and since this violation is achievable with quantum mechanics

[Mer90] and $Q \subseteq Q^{1+ABC}$, it is achievable in Q^{1+ABC} also. Therefore, ascertaining the maximal probability $P(\mathbf{a}|000)$ compatible with this violation and for correlations in Q^{1+ABC} is an optimization of the form:

$$\begin{aligned} &\text{maximize } P(\mathbf{a}|000) \\ &\text{subject to } \langle 001\rangle + \langle 010\rangle + \langle 100\rangle - \langle 111\rangle = 4, \\ &\qquad P(\mathbf{a}|000) \in Q^{1+ABC}. \end{aligned}$$

Given our construction of correlations in Q^{1+ABC}, we can rephrase this optimization in terms of a semidefinite program:

$$\begin{aligned} &\text{maximize } \frac{1}{2}\text{tr}(M\Gamma_{1+ABC}) \\ &\text{subject to } \frac{1}{2}\text{tr}(B\Gamma_{1+ABC}) = 4, \\ &\qquad \Gamma_{1+ABC} \succeq 0, \\ &\qquad \frac{1}{2}\text{tr}(D_i\Gamma_{1+ABC}) = 0, i \in \{1, 2, ..., m\}, \end{aligned}$$

where M, B and D_i are real, symmetric 15-by-15 matrices such that $\frac{1}{2}\text{tr}(M\Gamma_{1+ABC})$ $= P(\mathbf{a}|000)$ and $\frac{1}{2}\text{tr}(B\Gamma_{1+ABC}) = \langle 001\rangle + \langle 010\rangle + \langle 100\rangle - \langle 111\rangle$. Due to (D.1), we can impose the former equality on $\frac{1}{2}\text{tr}(M\Gamma_{1+ABC})$. The m matrices D_i just impose constraints on elements of Γ_{1+ABC} such that they are compatible with Q^{1+ABC}.

We now fix the particular representation of Γ_{1+ABC} with elements $[\Gamma_{1+ABC}]_{ij} = \langle\psi|\mathcal{O}_i\mathcal{O}_j|\psi\rangle$ such that for both rows i and columns j we write the ordered vector of operators $\mathcal{O}_i$ with i increasing from left to right:

$$\begin{aligned} (\mathcal{O}_1, ..., \mathcal{O}_{15}) = (&\mathbb{I}, A_0, A_1, B_0, B_1, C_0, C_1, A_0B_1C_1, A_1B_0C_1, \\ &A_1B_1C_0, A_0B_0C_0, A_1B_0C_0, A_0B_1C_0, A_0B_0C_1, A_1B_1C_1). \end{aligned} \tag{D.-4}$$

Immediately we observe that the diagonal elements of the matrix $[\Gamma_{1+ABC}]_{ii} = 1$ and thus the magnitude of all elements of the matrix $|[\Gamma_{1+ABC}]_{ij}| \leq 1$ are bounded if the matrix is positive semidefinite. For example, given this representation $B = C + C^{\text{T}}$ where $C = \left(\begin{smallmatrix} v \\ \tilde{0} \end{smallmatrix}\right)$ for $\tilde{0}$ being a 14-by-15 matrix of zeroes and $v = (0, 0, ..., 0, 1, 1, 1, -1)$.

There is a unique solution to the problem in (D.2) if instead of the probability distribution being in Q^{1+ABC} it is constrained to be in Q. As mentioned $Q \subseteq Q^{1+ABC}$, so we can write this solution as a matrix of the form Γ_{1+ABC}, and we call this solution matrix Γ_{M} and define it as follows:

Definition 1. The only solution matrix Γ_{M} to (D.2) that can be realized in quantum theory has elements

1. $[\Gamma_{\mathrm{M}}]_{ij} = 1$ if $i = j$, $[\Gamma_{\mathrm{M}}]_{ij} \in \{\langle 001\rangle, \langle 010\rangle, \langle 100\rangle\}$ and $[\Gamma_{\mathrm{M}}]_{ij} \in \{\langle\psi|\mathcal{P}(\mathcal{P}')^{\dagger}|\psi\rangle, \langle\psi|(\mathcal{P}')^{\dagger}\mathcal{P}|\psi\rangle\}$ where $\mathcal{P}, \mathcal{P}' \in \{A_0A_1, B_0B_1, C_0C_1\}$ and $\mathcal{P} \neq \mathcal{P}'$;

2. $[\Gamma_{\mathrm{M}}]_{ij} = -1$ if $[\Gamma_{\mathrm{M}}]_{ij} = \langle 111\rangle$ and $[\Gamma_{\mathrm{M}}]_{ij} \in \{\langle\psi|\mathcal{P}\mathcal{P}'|\psi\rangle, \langle\psi|\mathcal{P}'\mathcal{P}|\psi\rangle\}$ where $\mathcal{P}$, $\mathcal{P}' \in \{A_0A_1, B_0B_1, C_0C_1\}$ and $\mathcal{P} \neq \mathcal{P}'$;

3. $[\Gamma_{\mathrm{M}}]_{ij} = 0$ otherwise.

We now present the main theorem of this section.

Theorem 1. *The only possible solution matrix Γ_{1+ABC} to the semidefinite program in* (D.2) *is Γ_M.*

This immediately leads to the following corollary that is relevant for randomness certification. That is, since the solution to the semidefinite program in (D.2) is the quantum solution we inherit the result of Dhara et al [DPA13] that shows that we obtain three random bits if we maximally violate the Mermin inequality [Mer90]. We state this result more formally in the following corollary.

Corollary D.1. *The maximal value of the objective function $\frac{1}{2}tr(M\Gamma_{1+ABC}) = P(\boldsymbol{a}|000)$ in the semidefinite program* (D.2) *is equal to $\frac{1}{8}$ for all $\boldsymbol{a}$.*

Proof – First we observe that, as obtained from the definition of matrix Γ_{M}, $\langle 0\rangle = 0$ for every party's single-body correlator, and equally $\langle 00\rangle = 0$ for all two-body correlators between the three parties, and $\langle 000\rangle = 0$. Substituting these values into (D.1), we then obtain $P(\mathbf{a}|000) = \frac{1}{8}$ for all $\mathbf{a}$. Since Γ_{M} is the only possible solution to (D.2), this is the only possible probability distribution over $\mathbf{a}$. □

To prove theorem 1 we require two lemmas that will be introduced and proved in the sequel. The first lemma describes the structure of the feasible matrices Γ_{1+ABC}, i.e. the matrices that satisfy all of the constraints in (D.2). The second lemma just says that matrices Γ_{1+ABC} of this form are positive semidefinite if and only if they are equal to Γ_{M}. We now present and prove these lemmas. For simplicity we utilize the notation for the notation $\langle O\rangle = \langle\psi|O|\psi\rangle$ with $O \in \{A_j, B_j, C_j | j \in \{0, 1\}\}$.

Lemma D.2. *Matrices* Γ_{1+ABC} *are feasible (satisfy all constraints therein) for the semidefinite program* (D.2) *if and only if they are of the form:*

$$\Gamma_{1+ABC} = \begin{pmatrix} 1 & \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \\ \boldsymbol{q}_1^T & \mathbb{W} & \mathbb{X} & \mathbb{Y} \\ \boldsymbol{q}_2^T & \mathbb{X}^T & \mathbb{D} & \mathbb{O} \\ \boldsymbol{q}_3^T & \mathbb{Y}^T & \mathbb{O} & \mathbb{D} \end{pmatrix}, \tag{D.-5}$$

with

$$\begin{aligned}
\boldsymbol{q}_1 &= (\langle A_0\rangle, \langle A_1\rangle, \langle B_0\rangle, \langle B_1\rangle, \langle C_0\rangle, \langle C_1\rangle)\,, \\
\boldsymbol{q}_2 &= (0,0,0,0)\,, \\
\boldsymbol{q}_3 &= (1,1,1,-1)\,, \\
\mathbb{W} &= \begin{pmatrix} \mathbb{I} & \mathbb{C} & \mathbb{B} \\ \mathbb{C}^T & \mathbb{I} & \mathbb{A} \\ \mathbb{B}^T & \mathbb{A}^T & \mathbb{I} \end{pmatrix}, \\
\mathbb{X} &= \begin{pmatrix} -\langle A_1\rangle & \langle A_1\rangle & \langle A_1\rangle & \langle A_1\rangle \\ -\langle A_0\rangle & \langle A_0\rangle & \langle A_0\rangle & \langle A_0\rangle \\ \langle B_1\rangle & -\langle B_1\rangle & \langle B_1\rangle & \langle B_1\rangle \\ \langle B_0\rangle & -\langle B_0\rangle & \langle B_0\rangle & \langle B_0\rangle \\ \langle C_1\rangle & \langle C_1\rangle & -\langle C_1\rangle & \langle C_1\rangle \\ \langle C_0\rangle & \langle C_0\rangle & -\langle C_0\rangle & \langle C_0\rangle \end{pmatrix}, \\
\mathbb{Y} &= \begin{pmatrix} \langle A_0\rangle & \langle A_0\rangle & \langle A_0\rangle & -\langle A_0\rangle \\ \langle A_1\rangle & \langle A_1\rangle & \langle A_1\rangle & -\langle A_1\rangle \\ \langle B_0\rangle & \langle B_0\rangle & \langle B_0\rangle & -\langle B_0\rangle \\ \langle B_1\rangle & \langle B_1\rangle & \langle B_1\rangle & -\langle B_1\rangle \\ \langle C_0\rangle & \langle C_0\rangle & \langle C_0\rangle & -\langle C_0\rangle \\ \langle C_1\rangle & \langle C_1\rangle & \langle C_1\rangle & -\langle C_1\rangle \end{pmatrix}, \\
\mathbb{D} &= \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix},
\end{aligned}$$

$\mathbb{O}$ *being a* 4*-by-*4 *matrix of all-zeroes, with* $\mathbb{A} = \left(\begin{smallmatrix} \langle A_1\rangle & \langle A_0\rangle \\ \langle A_0\rangle & -\langle A_1\rangle \end{smallmatrix}\right)$, $\mathbb{B} = \left(\begin{smallmatrix} \langle B_1\rangle & \langle B_0\rangle \\ \langle B_0\rangle & -\langle B_1\rangle \end{smallmatrix}\right)$, $\mathbb{C} = \left(\begin{smallmatrix} \langle C_1\rangle & \langle C_0\rangle \\ \langle C_0\rangle & -\langle C_1\rangle \end{smallmatrix}\right)$ *and* $\mathbb{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Proof – Vectors $\mathbf{q}_1$ and $\mathbf{q}_3$ are trivially obtained if the constraints in (D.2) are satisfied.

We now use the observation that for all feasible matrices Γ_{1+ABC}, the elements $[\Gamma_{1+ABC}]_{ij} \in \{\langle 001\rangle, \langle 010\rangle, \langle 100\rangle\}$ are all equal to 1 and when $[\Gamma_{1+ABC}]_{ij} =$

$\langle 111\rangle$ the element is equal to -1. This is due to the fact that this is the only combination of values compatible with maximal violation of the Mermin inequality. This fact implies that $\langle\psi|\mathcal{R}|\psi\rangle = \langle\psi|\psi\rangle$ for $\mathcal{R} \in \{A_0B_0C_1, A_0B_1C_0, A_1B_0C_0\}$ and $\langle\psi|A_1B_1C_1|\psi\rangle = -\langle\psi|\psi\rangle$ and by normalization,

$$\begin{aligned} A_0B_0C_1|\psi\rangle &= |\psi\rangle, \\ A_0B_1C_0|\psi\rangle &= |\psi\rangle, \\ A_1B_0C_0|\psi\rangle &= |\psi\rangle, \\ A_1B_1C_1|\psi\rangle &= -|\psi\rangle. \end{aligned}$$

This implies that $\langle\psi|\mathcal{P}\mathcal{R}|\psi\rangle = \langle\psi|\mathcal{P}|\psi\rangle$ for $\mathcal{R} \in \{A_0B_0C_1, A_0B_1C_0, A_1B_0C_0\}$ and $\langle\psi|\mathcal{P}A_1B_1C_1|\psi\rangle = -\langle\psi|\mathcal{P}|\psi\rangle$ where $\mathcal{P}$ is any $\mathcal{O}_j$ as described above for $j \in \{1, 2, ..., 15\}$. Utilising this observation we obtain the sub-matrix $\mathbb{D}$ in (D.2) if $\mathcal{P}$ is equal to any of the $\mathcal{R}$ described above. Also for $\mathcal{P} \in \{A_i, B_j, C_k\}$ for all i, j, k, we again utilize this observation to obtain $\mathbb{Y}$ and certain elements of $\mathbb{X}$. The elements of $\mathbb{X}$ that are obtained via this observation are those where $\langle\psi|\mathcal{O}_i\mathcal{O}_j|\psi\rangle = \langle\psi|\mathcal{P}\mathcal{R}|\psi\rangle$ with $\mathcal{P}$ and $\mathcal{R}$ being as described above.

To obtain the remaining elements of $\mathbb{X}$ that do not satisfy the above condition, we utilize another consequence of the conditions of (D.2). That is, since $\mathcal{O}_i\mathcal{O}_i = \mathbb{I}$, any element of Γ_{1+ABC} equal to $\langle\psi|\mathcal{S}|\psi\rangle$ for $\mathcal{S} \in \{A_iB_j, A_iC_k, B_jC_k\}$ is equal to $\pm\langle\psi|\mathcal{S}'|\psi\rangle$ for $\mathcal{S}' \in \{A_i, B_j, C_k\}$ only if
$\mathcal{S}\mathcal{S}' \in \{A_0B_0C_1, A_0B_1C_0, A_1B_0C_0, A_1B_1C_1\}$. The sign in front of $\langle\psi|\mathcal{S}'|\psi\rangle$ is determined by the product $\mathcal{S}\mathcal{S}'$. We also use this observation to obtain matrices $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$.

It remains to be shown how the vector $\mathbf{q}_2$, the matrix $\mathbb{O}$ and the submatrices $\mathbb{I}$ in $\mathbb{W}$ are obtained. We first observe that $\mathbf{q}_2 = (w, x, y, z)$ where $w = \langle\psi|A_0B_1C_1|\psi\rangle$, $x = \langle\psi|A_1B_0C_1|\psi\rangle$, $y = \langle\psi|A_1B_1C_0|\psi\rangle$, and $z = \langle\psi|A_0B_0C_0|\psi\rangle$. Utilising the relations in (D.2), we obtain

$$\mathbb{O} = \begin{pmatrix} w & w & w & -w \\ x & x & x & -x \\ y & y & y & -y \\ z & z & z & -z \end{pmatrix}. \tag{D.-15}$$

We now observe that $\mathbb{O}$ can be defined in an equivalent way since $\mathcal{O}_i\mathcal{O}_i = \mathbb{I}$ for all $\mathcal{O}_i$. Using this observation and $\langle\psi|\mathcal{O}_i\mathcal{O}_j|\psi\rangle = \langle\psi|\mathcal{O}_j\mathcal{O}_i|\psi\rangle$ for $\mathcal{O}_i$, $\mathcal{O}_j \in \{A_i, B_j, C_k\}$ and $\mathcal{O}_i \neq \mathcal{O}_j$, we obtain

$$\mathbb{O} = \begin{pmatrix} w & \langle C_0C_1\rangle & \langle B_0B_1\rangle & \langle A_0A_1\rangle \\ \langle C_0C_1\rangle & x & \langle A_0A_1\rangle & \langle B_0B_1\rangle \\ \langle B_0B_1\rangle & \langle A_0A_1\rangle & y & \langle C_0C_1\rangle \\ \langle A_0A_1\rangle & \langle B_0B_1\rangle & \langle C_0C_1\rangle & -z \end{pmatrix}, \tag{D.-15}$$

where again we are using the notation $\langle\psi|\mathcal{O}_i\mathcal{O}_j|\psi\rangle = \langle\mathcal{O}_i\mathcal{O}_j\rangle$ for brevity. Since the matrix in (D.2) and (D.2) have to be equal to each other, the only possible solution is that $\mathbb{O}$ is a 4-by-4 matrix of zeroes. This also implies that $\mathbf{q}_2 = (0,0,0,0)$ and $\langle A_0A_1\rangle = \langle B_0B_1\rangle = \langle C_0C_1\rangle = 0$, thus completing the matrix $\mathbb{W}$. This also completes our proof. □

We now present our final lemma that will complete the proof of theorem 1.

Lemma D.3. *The matrix* Γ_{1+ABC} *described by* (D.2) *is positive semidefinite if and only if* $\Gamma_{1+ABC} = \Gamma_M$.

Proof – We can use the Schur complement of Γ_{1+ABC} in (D.2) and that $\mathbb{D} = \mathbf{q}_3^T \cdot \mathbf{q}_3$ and $\mathbb{Y} = \mathbf{q}_1^T \cdot \mathbf{q}_3$ to show that Γ_{1+ABC} is positive semidefinite if and only if

$$\begin{pmatrix} \mathbb{W}' & \mathbb{X} \\ \mathbb{X}^T & \mathbb{D} \end{pmatrix} \succeq 0, \tag{D.-15}$$

where $\mathbb{W}' = \mathbb{W} - \mathbf{q}_1^T \cdot \mathbf{q}_1$. For example, for the matrix Γ_{M}, the corresponding submatrix Γ'_{M} from (D.2) is

$$\Gamma'_{\mathrm{M}} = \begin{pmatrix} \mathbb{I} & \overline{0} \\ \overline{0}^T & \mathbb{D} \end{pmatrix}, \tag{D.-15}$$

where $\mathbb{I}$ is the 6-by-6 identity matrix and $\overline{0}$ is a 6-by-4 matrix of zeroes. This submatrix of Γ_{M} is positive semidefinite if and only if $\mathbb{D} \succeq 0$ which is indeed true.

Since the space of positive semi-definite matrices is convex, the set of feasible matrices Γ_{1+ABC} for the semidefinite program (D.2) is a convex set. Therefore, if there is a submatrix Γ_1 of the form (D.2), we can obtain another submatrix Γ_2 of the form (D.2) that is a convex combination of Γ_1 and Γ'_{M}. We assume that Γ_1 has elements corresponding to some non-zero values $\{\langle A_i\rangle, \langle B_j\rangle, \langle C_k\rangle\}$, therefore completely unlike Γ_{M}. We now show that there exist matrices of the form Γ_2 that are not positive semidefinite which implies that any matrix Γ_1 as described is not positive semidefinite. This in turn implies that the only positive semidefinite matrix of the form (D.2) is Γ'_{M}.

We choose Γ_2 such that $\sum_{j=0}^{1} |\langle A_j\rangle| + |\langle B_j\rangle| + |\langle C_j\rangle| \ll 1$ but at least one of the elements of the set $\{\langle A_i\rangle, \langle B_j\rangle, \langle C_k\rangle\}$ is non-zero. As mentioned before, since the space of solution matrices Γ_{1+ABC} is convex we can always choose such a matrix without loss of generality. Therefore, the matrix in (D.2) is positive semidefinite if and only if

$$\begin{pmatrix} \overline{\mathbb{W}'} & \overline{\mathbb{X}} \\ \overline{\mathbb{X}}^T & \mathbb{D} \end{pmatrix} - \frac{1}{(1 - \langle A_0\rangle^2)} \begin{pmatrix} \mathbf{s}^T \\ \mathbf{r}^T \end{pmatrix} \cdot \begin{pmatrix} \mathbf{s} & \mathbf{r} \end{pmatrix} \succeq 0, \tag{D.-15}$$

where $\mathbb{X} = \left(\frac{\mathbf{r}}{\overline{\mathbb{X}}}\right)$ where $\mathbf{r} = (-\langle A_1\rangle, \langle A_1\rangle, \langle A_1\rangle, \langle A_1\rangle)$ is the first row of $\mathbb{X}$, $\overline{\mathbb{W}'}$ is $\mathbb{W}'$ without the first column and first row, and $\mathbf{s}$ is the first row of $\mathbb{W}'$ excluding the element $[\mathbb{W}']_{11}$. Since every diagonal element of $\overline{\mathbb{W}'} - \frac{1}{(1-\langle A_0\rangle^2)}\mathbf{s}^T \cdot \mathbf{s}$ is positive by construction, then the matrix in (D.2) is positive semidefinite if and only if $\mathbb{E} = \mathbb{D} - \frac{1}{(1-\langle A_0\rangle^2)}\mathbf{r}^T \cdot \mathbf{r} \succeq 0$. Note that every diagonal element of $\mathbb{E}$ is equal to $1 - \frac{1}{(1-\langle A_0\rangle^2)}\langle A_1\rangle^2$. However, the element $[\mathbb{E}]_{12} = 1 + \frac{1}{(1-\langle A_0\rangle^2)}\langle A_1\rangle^2$. For a matrix to be positive semedefinite off-diagonal elements have a magnitude that is bounded by the diagonal terms, therefore for $\mathbb{E}$ to be positive semidefinite we must satisfy $\langle A_1\rangle = 0$.

We can now repeatedly apply the same analysis to subsequent bottom-left submatrices of (D.2) where the matrix in (D.2) is positive semidefinite if and only if $\mathbb{E}' = \mathbb{D} - \frac{1}{\alpha}\mathbf{r'}^T \cdot \mathbf{r'} \succeq 0$ where $\alpha < 1$ is some positive real number and $\mathbf{r'}$ is any row of $\mathbb{X}$. For every matrix $\mathbb{E}'$ the diagonal elements are $1 - \frac{1}{\alpha}\langle\mathcal{P}\rangle^2$ where $\mathcal{P} \in \{A_i, B_j, C_k\}$ but there are off-diagonal terms in $\mathbb{E}'$ that take the value $1 + \frac{1}{\alpha}\langle\mathcal{P}\rangle^2$. Therefore for all Γ_{1+ABC} described by (D.2), $\langle\mathcal{P}\rangle = 0$ for $\mathcal{P} \in \{A_i, B_j, C_k\}$ for all i, j, k. This matrix thus corresponds to Γ_{M} and completes our proof. □

Combining the two lemmas above we then obtain our proof of Theorem 1. This concludes our observation that maximally random numbers can be certified within a set of correlations that is not the quantum set. Our proof is analytic and makes concrete the numerical observations in Dhara et al [DPA13]. It would be interesting to extend this proof to other scenarios even though we have used a lot of the structure of the Mermin inequality and the $(3, 2, 2)$ scenario.

E. Proofs for the protocols of chapter 8

E.1. Distinguishing two computable preparations of the same mixed quantum state

In this section we discuss with details the protocol outlined in chapter 8 to distinguish two ensembles of pure states that apparently yield the same density matrix, given the assumption that the mixture has been prepared in a computable way. To do so, we will first reduce this scenario to a different problem in classical information theory relating infinite binary sequences, which we show to be solvable. Then we see how this second problem can be solved in finite time with arbitrarily small error probability. Finally, we present a slight modification of the algorithm that makes it robust against a simple noise model.

E.1.1. The distinction protocol

The distinguishability scenario we are interested in is as follows: Alice is presented with two bags of quantum systems, one having systems in the state $|0\rangle$ and the other having systems in the state $|1\rangle$. At random, she chooses one bag at a time and sends a state from that bag to Bob. Bob's state is $\rho = \mathbb{I}/2$, as he has no information about the prepared state. Now imagine the same scenario, but instead of those two bags Alice has one with the state $|+\rangle$ and another with the state $|-\rangle$, and she proceeds in the same way. It is clear that these two situations are indistinguishable from Bob's point of view, as they are supposed to define the same mixed quantum state.

Consider the same two situations, but now, in order to choose from which bag to pick the state, Alice uses the bits of a computable binary sequence, that is, a binary sequence produced by an algorithm unknown to Bob. Ideally such algorithm is a pseudo-random number generator, whose output 'looks like' typical coin tossing (i.e. satisfying for instance the law of large numbers, as well as any other reasonable law of randomness [LV08]). We next see that, if Alice's sequence is computable, Bob can distinguish between both situations in finite time and with an arbitrarily high probability of success.

Bob's protocol works as follows: he measures Alice's first qubit in the basis of eigenstates of σ_x, the second qubit in the basis of eigenstates of σ_z, and so on, measuring every odd qubit in the σ_x basis and every even qubit in the σ_z basis. The output of the measurements would yield in the limit two infinite binary sequences: X, obtained from the measurements on the odd qubits and Z, obtained from the even qubits. Now X and Z have a distinctive feature: when Bob measures in the same basis as Alice prepared the states, the sequence obtained is computable (because it is either the odd or the even bits of a computable sequence), and when the measurement is performed in the other basis the sequence is similar to one obtained from tossing a fair coin. Thus, if Bob can distinguish a computable sequence from a fair coin he can tell what was the basis in which Alice prepared the state. We will see that this is indeed the case with arbitrarily high probability of success. Both preparations are, thus, distinguishable. It is important to remark that, since Bob will only need finite prefixes from both sequences to achieve the distinction, he just needs to measure a finite number of qubits received from Alice.

As mentioned, after his measurements, Bob is left with the problem of distinguishing a computable sequence from a fair coin tossing. We present an algorithm that can do this with an arbitrarily high probability of success: given two sequences X and Z and a desired error probability e, our algorithm decides in finite time (that is, after checking finitely many bits) which of the two sequences is the computable one, giving a wrong answer with a probability smaller or equal than e. The hand-wavy idea of the algorithm is to check every program (the set of Turing machines is enumerable) on an universal Turing machine U until finding one that reproduces a sufficiently long prefix of either X or Z, say X. It is then claimed that X is the computable sequence, that is, the basis used by Alice for encoding. We will now explore what subtleties appear when trying to follow that idea.

E.1.2. Distinguishing a fair coin from a computer

In this Section we show the algorithm that can distinguish, with arbitrarily small error probability, a computable sequence from one arising from a fair coin.

Notation and model of computation

Let us first fix some notation. The set of finite strings over the alphabet $\{0,1\}$ is denoted 2^* and ϵ denotes the empty string. The set of infinite sequences over $\{0,1\}$ is denoted 2^ω. If $S \in 2^\omega$ then $S \upharpoonright n$ denotes the string in 2^* formed by

the first n symbols of S. If $x, y \in 2^*$ then $x \preceq y$ represents that x is a prefix of y.

For the sake of completeness let us quickly restate the particular model of computation that we use for our results. We introduced this model in the preliminaries chapter 2.4.1 emphasizing that it is equivalent to that of regular Turing machines and thus having their same computational power. Specifically, we consider Turing machines M with a reading, a working and an output tape (the last two being initially blank). The output of M for input $x \in 2^*$ is denoted $M(x) \in 2^*$, and if $t \geq 0$, $M_t(x) \in 2^*$ consists of the content of the output tape in the execution of M for input x by step t —notice that this execution needs not be terminal, that is, $M(x)$ needs not be in a halting state at stage t. A *monotone* Turing machine is a Turing machine whose output tape is one-way and write-only, meaning that it can append new bits to the output but it cannot erase previously written ones. Hence if M is a monotone machine $M_t(x) \preceq M_s(x)$ for $t \leq s$.

A sequence $S \in 2^\omega$ is *computable* if there is a (monotone) Turing machine M such that for all n, $M(n) = S \upharpoonright n$. Equivalently, S is computable if there is a monotone machine M such that $M(\epsilon)$ "outputs" S, in the sense that

$$(\forall n)(\exists t)\ M_t(\epsilon) = S \upharpoonright n. \tag{E.0}$$

Let $(M_i)_{i\geq 0}$ be an enumeration of all monotone Turing machines and let U be a monotone Turing machine defined by $U(\langle i, x\rangle) = M_i(x)$, where $\langle \cdot, \cdot \rangle : \mathbb{N}^2 \to \mathbb{N}$ is any computable pairing function (i.e. one that codifies two numbers in $\mathbb{N}$ into one and such that both the coding and the decoding functions are computable). The machine U is universal for the class of all monotone machines. In other words, U is an interpreter for the class of all monotone Turing machines, and the argument p in $U(p)$ is said to be a *program* for U, encoding a monotone Turing machine and an input for it.

In the following subsection we deal with a related problem: distinguishing a computable sequence from the output of a 'fair coin' (such as the result of measuring a σ_z eigenstate in the σ_x basis, under the assumption that quantum physics is correct). Notice that since there are countably many computable sequences, the output of tossing a fair coin gives a non-computable sequence with probability one. We show that one can distinguish both cases in finite time and with arbitrarily high success probability and that this fact has consequences for how mixed states in quantum mechanics are described.

The protocol

As mentioned, the idea of the algorithm is to check every program until finding one that reproduces a sufficiently long prefix of either X or Z. There are three key points that have to be taken into account in this idea, namely:

- It is impossible to know if a program halts or not [Tur36]. Therefore, checking each single program one after another is not possible.
- It is still not clear what we mean by *sufficiently long* prefix.
- We might get a false positive, i.e. find a program that reproduces a prefix of the sequence which came form the coin tossing (even if it was not computable).

We deal with the first issue by dovetailing between programs and execution time. Recall that programs for U can be coded by (binary representations of) natural numbers. The idea of dovetailing is that we first run program 0 for 0 steps, then we run programs 0 and 1 for 1 step, then we run programs 0, 1 and 2 for 2 steps and so on.

To solve the second problem, we check if program p of length $|p|$ generates (within the time imposed by the dovetailing) either of the prefixes $X \upharpoonright j$ (that is, the first j bits of X) or $Z \upharpoonright j$, where $j = k|p|$. That is, every program is checked against a prefix k times longer than its length. Since a fair coin generates sequences with mostly non-compressible prefixes, it most likely will not have a prefix that can be generated by a k times shorter program, thus allowing us to detect the computable sequence. And as we will see, the probability of getting a false positive can be bounded only by a function of k that goes to 0 as k goes to infinite, solving also the third issue.

The pseudo-code for the algorithm that decides which of the sequences is computable will be the following:

Note that X and Z are infinite sequences, and hence they must be understood as *oracles* [Soa87, SIII] in the effective procedure described above. Provided that at least one of X or Z is computable, the above procedure always halt —and so it only queries finitely many bits of both X and Z. Indeed, in case $S \in \{X, Z\}$ is computable, there is a monotone Turing machine $M = M_i$ and $i \in \mathbb{N}$ such that $M(\epsilon)$ outputs S in the sense of (E.1.2). Hence for program $p = \langle i, \epsilon \rangle$ we have that $U_t(p) = S \upharpoonright k|p|$ for some t.

It is important to recall that although both X and Z are infinite sequences, we only need to query finite prefixes. From a physical point of view this means that only finitely many qubits will be needed by Bob to discover how Alice was preparing the state.

Algorithm 3 The distinguishing protocol

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

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

Now, we bound the probability of having a miss-recognition, that is, the probability P_{error} that the above procedure outputs 'Z' when X was computable, or viceversa. To do so, we bound the probability that $S \in 2^\omega$ has the property that for the given value of k there is p such that

$$(\exists t)\ U_t(p) = S \upharpoonright k|p|. \tag{E.0}$$

Since there are 2^ℓ programs of length ℓ, the probability that there is a program p of length ℓ such that (E.1.2) holds is at most $2^\ell/2^{k\ell}$. Adding up over all possible lengths ℓ we obtain

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

which goes to zero exponentailly with k.

The protocol would then work as follows: given a tolerated error probability e, one chooses a k large enough so that the previous bound is smaller than e, and then run the described algorithm with inputs k, X and Z.

E.1.3. Noise robustness

We now show how to modify the previous algorithm to make it robust against noise. We can consider a very natural noise model in which random bit flips are applied to the measured sequences, resulting for instance from imperfect preparations or measurements. Therefore, we modify the algorithm so that it

Algorithm 4 The noise tolerant distinguishing protocol

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

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

tolerates a fraction $q \in \mathbb{Q}$ of bit flips in the prefixes. The modified algorithm is as follows:

where d_H is the Hamming distance between two strings, which counts the number of different bits in both strings. The first thing to notice is that when $q = 0$ Algorithms 3 and 4 coincide.

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

$$P_{\text{error}} \leq \sum_{\ell > 0} \frac{2^\ell 2^{\lfloor q\ell k \rfloor} \binom{\ell k}{\lfloor q\ell k \rfloor}}{2^{\ell k}} \tag{E.0}$$

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

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

This geometric sum can be easily computed yielding

$$P_{\text{error}} \leq \frac{2^{1+qk-k} \left(\frac{e}{q}\right)^{qk}}{1 - 2^{1+qk-k} \left(\frac{e}{q}\right)^{qk}}. \tag{E.0}$$

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

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

E.2. The Bell test computability loophole

In this section we discuss how the knowledge that measurement settings are chosen using a device that satisfies the Church-Turing thesis opens a new loophole in Bell tests. We focus our attention on the simplest Bell scenario although generalizations are straightforward: let us consider a bipartite scenario in which the two parties, Alice and Bob, have a box each with two input buttons (left and right) and a binary output. A source between them is sending physical systems sequentially to each party. Upon arrival, Alice and Bob choose what input buttons to press thus performing different measurements on the particles. Our object of interest is the probability distribution describing the process $P(a, b|x, y)$ where x and y are inputs for Alice and Bob's box respectively and a and b are their outputs which can be derived from the statistics.

Let us also imagine that the input-output events at each site define space-like separated events so that Bob's input cannot influence Alice's output and viceversa. Focusing on a particular round of the experiment, let us describe by λ a complete set of variables (some of which could be hidden or unknown) that characterize the physical systems such that the outcomes of the measurements are deterministic $P(a, b|x, y, \lambda) = \delta^a_{f(x,\lambda)} \delta^b_{g(y,\lambda)}$ where the functions $f(x, \lambda)$ and

$g(y, \lambda)$ map determinstically inputs x, y to outputs a, b, using the complete description λ. Different rounds of the experiment may be described with different set of variables λ drawing from a probability distribution $p(\lambda|x, y) \equiv p(\lambda)$ where the equality condition is termed *measurement independence* or *free choice*. This crucial condition implies that the complete description λ is independent from the choice of settings x, y that Alice and Bob will use to measure the systems. Thus, we say that a probability distribution is *local* if it can be written as

$$P(a, b|x, y) = \int p(\lambda)\delta^a_{f(x,\lambda)}\delta^b_{g(y,\lambda)}d\lambda. \tag{E.0}$$

It can be shown that any probability distribution that violates the following Bell inequality, namely a CHSH inequality [CHSH69], is not a local distribution:

$$\sum_{a,b,x,y\in\{0,1\}} (-1)^{a+b+x\cdot y} P(a, b|x, y) \leq 2. \tag{E.0}$$

Remarkably, quantum correlations, obtained for example by measuring a maximally entangled two-qubit state with non-commuting measurements, can violate this inequality. The violation of a Bell inequality witnesses the existence of non-local correlations which in turn can be used in many device-independent applications such as randomness expansion or for establishing a secure key between distant locations. Hence, checking whether the experimental data truly violates a Bell Inequality is of outmost importance for device-independent information science as we have greatly emphasized throughout this thesis.

In order to present the computability loophole, we introduce an eavesdropper named Eve. Eve will be able to prepare Alice's and Bob's local boxes at the beginning of the experiment as shown in Fig. 8.3. Alice's box will then have access to the inputs of both parties after the measurements are done (Alice's input in a straightforward way, and Bob's input via classical communication). This scenario was first termed the two-sided memory loophole in the literature [BCH$^+$02]. An equivalent scenario has been used more recently in [PAM$^+$10c, PM13] so as to perform device-independent randomness expansion where they allow the boxes behaviour to adapt depending on the information of previous rounds. Interestingly, local models exploiting this past information have been shown to be of no help to violate the Bell Inequality in the asymptotic limit. Indeed, the probability that a local model reproduces some observed violation despite using past inputs and outputs goes exponentially fast to zero in the number of rounds [PM13].

As mentioned earlier, a crucial condition for Bell tests to establish nonlocality and randomness is to assume measurement independence $p(\lambda|x, y) \equiv p(\lambda)$,

which in our context is independence between the boxes prepared by Eve and the measurement choices of Alice and Bob. Let us imagine that Alice and Bob chose their measurement settings following an algorithm which is a standard practice in all Bell experiments to date. Trivially, if Alice's box knows which algorithms Alice and Bob are using, it can fake a Bell violation. We assume that the algorithms used by Alice and Bob are fully unknown to each other and to Eve, thus uncorrelated to the boxes she initially prepared.

Our result is to show that if either Alice or Bob (or both) choose their measurements following an algorithm —or equivalently, because of the Church-Turing thesis, following any classical mechanical procedure—, even under the assumption that such algorithm is fully unknown to Eve and hence uncorrelated to the boxes she initially prepares, there is an attack that, in the asymptotic limit, produces a Bell inequality violation from purely deterministic boxes, thus providing the aforementioned loophole.

Before proceeding, we need a few more tools from computer science. We say that a class $\mathcal{C}$ of computable functions is a *time [resp. space] complexity class* if there is a computable function t such that each function in $\mathcal{C}$ is computed by a Turing machine that, for every input x, runs in time [resp. space] $O(t(|x|))$. Examples of such classes include the well-known **P**, **BQP**, **NP**, **PSPACE** where the complexity time bound is a simple exponential function and the much broader class **PR** of the primitive recursive functions where the time bound is Ackermannian [Odi99, SVIII.8] (see [Kup] for an inclusion diagram of the most well-studied complexity classes).

Let us now say that one party, say Alice without loss of generality, will be using an algorithm to produce her measurements choices. In formal terms, this means that there is a computable function $f_A : \mathbb{N} \to \{\text{left}, \text{right}\}$ such that $f_A(i)$ tells Alice to press the left or the right button at the i-th round.

As we previously pointed out, it is clear that if Eve knows (any algorithm for) f_A, her task becomes trivial. In our setting, however, function f_A is unknown to Eve when she prepared the boxes. However, we assume the following further hypothesis: Eve knows *some* time or space bound t of a complexity class containing f_A and f_B (the corresponding function for Bob's inputs). For instance, Eve knows that Alice and Bob use at most, say, time $O(t(n))$, for $t(n) = 2^{2^n}$ (though the algorithms that Alice is actually running may take, say, $O(n^2)$). It is important to note that this hypothesis is quite mild, because every computable function belongs to some time or space complexity class —given a program there is a computable interpreter which executes it for some given input by stages and counts the number of steps that such execution takes to terminate or the number of cells used in the tape. In other words, for every computable function g there is a computable function t_g that upper bounds the

running time or space of some algorithm for g.

Knowing this time or space bound t, Eve can program a computing device in one of the boxes, say Alice's, to *predict* the functions f_A and f_B from some point onwards. This means that Alice's box has an effective procedure that, after having seen $f_A(0), f_A(1), \ldots, f_A(k)$ for large enough k, allows her to correctly guess $f_A(k+1), f_A(k+2)\ldots$ and the same for f_B. The existence of such k will be guaranteed by Alice's box procedure; however it will not be able to effectively determine when this k has arrived. The idea behind this is that every time bounded class is computably enumerable, allowing Eve to pick, at each round, the first program for a function from that class that reproduces the inputs given by Alice (or Bob) so far. Since the function used by Alice (or Bob) belongs to that class, at some point the first program that Alice will find reproducing the inputs given so far will be one which computes the function used by Alice (or Bob), therefore allowing Alice's box to predict every input to come. See Sec. E.2.1 for a detail of Alice's box procedure.

Back to the loophole, under these assumptions, Eve is able to prepare both boxes so as to fake a Bell Inequality violation. Moreover, she could even prepare boxes that seem to be more non-local than what quantum mechanics allows. To see how, notice that any bipartite no-signaling probability distribution, local or not, can always be written as

$$P(a,b|x,y) = \int p(\lambda)\delta^a_{f(x,\lambda)}\delta^b_{g(y,x,\lambda)}d\lambda \tag{E.1}$$

$$= \int p'(\lambda)\delta^b_{f'(y,\lambda)}\delta^a_{g'(x,y,\lambda)}d\lambda \tag{E.2}$$

where again functions f, f', g, g' are deterministic functions (See Sec. E.2.1 for a prove). This means that, given that Eve learns either Alice's input x or Bob's input y, she can prepare deterministic (local) boxes to simulate any probability distribution and hence fake any Bell Inequality violation.

E.2.1. Predicting computable functions from initial segments

The theory of predicting computable functions started with the seminal works by Solomonoff [Sol64a, Sol64b] on inductive inference, and Gold [Gol67] on learnability. It studies the process of coming up with, either explanations (in the form of computer programs) or next-value predictions, after seeing some sufficiently big subset of the graph of a computable function. Many possible formalizations, depending on how the data is presented and how the learning process converges, have been considered in the literature (see [ZZ08] for a comprehensive survey). The most suitable model for our purposes is called *identifi-*

cation by next value, and follows by elementary arguments from computability theory.

A class of total computable functions $\mathcal{C}$ is *identifiable by next value* ($\mathcal{C} \in \mathrm{NV}$) [Bar71] if there exists a computable function g (called a *next-value function for* $\mathcal{C}$) such that for every $f \in \mathcal{C}$,

$$(\exists n_0)(\forall n \geq n_0)\ f(n) = g(\langle f(0), \ldots, f(n-1)\rangle). \tag{E.2}$$

Here $\langle x_1, \ldots, x_n\rangle$ is any computable codification of an n-tuple with a natural number, whose decoding is also computable. Condition (E.2.1) formalizes the idea that given the past values of f (namely $(f(0), \ldots, f(n-1))$), g can predict the forthcoming value of f (namely, $f(n)$), provided n is large enough —how large depends on the function f that we want to learn.

It follows from a simple diagonal argument that the class of all computable functions is not in NV. However, any time or space complexity class is NV. Indeed, suppose $\mathcal{C}$ is a time complexity class with (computable) time bound t. The following algorithm computes a next-value function for $\mathcal{C}$. Let $(M_i)_{i\in\mathbb{N}}$ be an enumeration of all Turing machines.

Algorithm 5 A next-value algorithm for a time class $\mathcal{C}$ with computable bound t

Require: $n \in \mathbb{N}$

Ensure: $g(n)$, the next-value function for $\mathcal{C}$.

Let M_e be an enumeration of all Turing machines.

Let $n = \langle m_0, \ldots, m_{n-1}\rangle$ be the already seen bits from the sequence.

Let $\langle e, c\rangle$ be the least number such that

i. for $i \in \{0, \ldots, n\}$, $M_e(i)$ halts after at most $c \cdot t(|i|)$ many steps, where t is the computable time bound for class $\mathcal{C}$.

ii. for $i \in \{0, \ldots, n-1\}$ $M_e(i)$ outputs m_i

Output $M_e(n)$

Suppose $f \in \mathcal{C}$, i.e. there is some Turing machine $M_{e'}$ and constant c' such that for every $x \in \mathbb{N}$,

$$M_{e'}(x) \text{ computes } f(x) \text{ with time bound } c' \cdot t(|x|). \tag{E.2}$$

Both e' and c' are unknown, and the idea of Algorithm 5 is to try different candidates e and c for e' and c' respectively, until one is found. On input $n = \langle f(0), \ldots, f(n-1)\rangle$ the algorithm proposes a *candidate* Turing machine M_e which 'looks like' f on $0, \ldots, n-1$, and then guesses that $f(n)$ is the value

computed by $M_e(n)$. To be a candidate means not only to compute the same first n values, but also to do it within the time bound imposed by $\mathcal{C}$, which is $c \cdot t$. Of course, the chosen candidate may be incorrect because, for instance, for input $\langle f(0), \ldots, f(n-1), f(n)\rangle$ we may realize that $f(n)$ was not equal to $M_e(n)$. In this case, the algorithm changes its mind and proposes as candidate a new pair $\langle e, c\rangle$. The existence of the correct candidates e' and c' satisfying (E.2.1) guarantees that:

1. For each n and each input $\langle f(0), \ldots, f(n-1)\rangle$ the algorithm will find some $\langle e, c\rangle$ meeting conditions i and ii.

2. Along the initial segments $\langle f(0), \ldots, f(n-1)\rangle$ for larger and larger n there can only be finitely many mind changes. Indeed, if the number of mind changes were infinite, then $\langle e', c'\rangle$ would be ruled out and this is impossible, as conditions i and ii are true for $e = e'$ and $c = c'$.

Hence there is n_0 such that for all $n \geq n_0$, the algorithm makes no more mind changes, and it stabilizes with values $\langle e, c\rangle$, which may not necessarily be equal to $\langle e', c'\rangle$, but will satisfy that $M_e(x)$ computes $f(x)$ with time bound $c \cdot t(|x|)$. Thus for input $\langle f(0), \ldots, f(n-1)\rangle$ the algorithm will return $f(n)$, and hence (E.2.1) will be satisfied. Observe that although the algorithm starts correctly predicting f from one point onwards, it cannot detect when this begin to happen. In other words, n_0 is not uniformly computable from e' and c'.

The algorithm for a space complexity class with bound t is analogous, but condition i must be modified to

i'. for $i \in \{0, \ldots, n\}$, $M_e(i)$ halts after at most $2^{t(|i|)}$ many steps and uses at most $t(|i|)$ many cells of the work tape during its computation.

Observe that any halting computation which consumes $t(n)$ many cells of the work tape runs for at most $2^{t(n)}$ many steps, as this is the total number of possible memory configurations. In condition i' we add the statement about the number of steps in order to avoid those computations which use at most $t(n)$ many cells but are non-terminating.

Simulation of no-signaling correlations from deterministic boxes

For completeness, we give a simple proof of the well-known fact that, if the input of one party in a Bell test is known, one can simulate any no-signaling distribution by using deterministic boxes. Let us imagine without loss of generality that it is Bob's box the one that has access to Alice's input x. First, notice that any no-signaling box can be written in the following way

$$P(a,b|x,y) = P(a|x)P(b|y;a,x) \equiv P(a|x)P_{a,x}(b|y). \tag{E.2}$$

Trivially, any local distribution can be simulated through deterministic boxes as $P(a|x) = \int p(\lambda)\delta^a_{f(x,\lambda)}d\lambda$. Hence, the no-signaling bipartite distribution can be written as

$$P(a,b|x,y) = \int\int p(\lambda)p'(\lambda')\delta^a_{f(x,\lambda)}\delta^b_{g(y,\lambda',a,x)}d\lambda d\lambda' \tag{E.2}$$

by defining now $\lambda'' = (\lambda,\lambda')$ and therefore $d\lambda'' = d\lambda d\lambda'$ and $p'(\lambda'') = p(\lambda)p'(\lambda')$ we have

$$P(a,b|x,y) = \int p''(\lambda)\delta^a_{f(x,\lambda'')}\delta^b_{g(y,\lambda'',x)}d\lambda''. \tag{E.2}$$

Notice that since a is a deterministic function of x and λ, given that λ'' includes the information of λ, the function on Bob's side does not need to depend explicitly on a.

Bibliography

[AAC+10] A. Acín, R. Augusiak, D. Cavalcanti, C. Hadley, J. K. Korbicz, M. Lewenstein, Ll, and M. Piani. Unified Framework for Correlations in Terms of Local Quantum Observables. *Physical Review Letters*, 104(14):140404+, April 2010.

[AB09] Elias Amselem and Mohamed Bourennane. Experimental four-qubit bound entanglement. *Nature Physics*, 5(10):748–752, 2009.

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

[ABG+07] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98(23):230501–, June 2007.

[AD12] Pablo Arrighi and Gilles Dowek. The physical Church-Turing thesis and the principles of quantum theory. *Int. J. Found. Comput. Sci.*, 23(5):1131–1146, 2012.

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

[AGCA12] Leandro Aolita, Rodrigo Gallego, Adán Cabello, and Antonio Acín. Fully nonlocal, monogamous, and random genuinely multipartite quantum correlations. *Phys. Rev. Lett.*, 108(10):100401–, March 2012.

[AGT06] Antonio Acín, Nicolas Gisin, and Benjamin Toner. Grothendieck's constant and local models for noisy entangled quantum states. *Phys. Rev. A*, 73:062105, Jun 2006.

[AMP12a] Antonio Acín, Serge Massar, and Stefano Pironio. Randomness versus nonlocality and entanglement. *Phys. Rev. Lett.*, 108(10):100402–, March 2012.

[AMP12b] Antonio Acín, Serge Massar, and Stefano Pironio. Randomness versus nonlocality and entanglement. *Phys. Rev. Lett.*, 108:100402, Mar 2012.

[Bak02] Andrew Baker. *Matrix groups : an introduction to Lie group theory*. Springer undergraduate mathematics series. Springer, London, Berlin, New York, 2002.

[Bar71] J.M. Barzdin. Prognostication of automata and functions. In *Information Processing '71*, pages 81–84, 1971.

[Bar07] Jonathan Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75(3):032304–21, March 2007.

[BBB+10] H. Barnum, S. Beigi, S. Boixo, M. B. Elliott, and S. Wehner. Local quantum measurement and no-signaling imply quantum correlations. *Phys. Rev. Lett.*, 104:140401, Apr 2010.

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

[BBLW07] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Generalized No-Broadcasting Theorem. *Physical Review Letters*, 99:240501+, December 2007.

[BBS+13] Julio T. Barreiro, Jean-Daniel Bancal, Philipp Schindler, Daniel Nigg, Markus Hennrich, Thomas Monz, Nicolas Gisin, and Rainer Blatt. Demonstration of genuine multipartite entanglement with device-independent witnesses. *Nat Phys*, 9(9):559–562, 09 2013.

[BBT11] Jop Briët, Harry Buhrman, and Ben Toner. A generalized grothendieck inequality and nonlocal correlations that require high entanglement. *Communications in Mathematical Physics*, 305(3):827–843, 2011.

[BC90] S.L. Braunstein and C.M. Caves. Wringing out better bell inequalities. *Annals of Physics*, 202:22–, 1990.

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

[BCK10] Arthur T Benjamin, Bob Chen, and Kimberly Kindred. Sums of evenly spaced binomial coefficients. *Mathematics Magazine*, 83(5):370–373, 2010.

[BCMdW10] Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Rev. Mod. Phys.*, 82(1):665–, March 2010.

[BCP+14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419–478, Apr 2014.

[BdlTS+14] Ariel Bendersky, Gonzalo de la Torre, Gabriel Senno, Santiago Figueira, and Antonio Acin. Implications of computer science principles for quantum physics. *arXiv preprint arXiv:1407.0604*, 2014.

[BdOFV10] Jop Briët, FernandoMário de Oliveira Filho, and Frank Vallentin. The positive semidefinite grothendieck problem with rank constraint. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and PaulG. Spirakis, editors, *Automata, Languages and Programming*, volume 6198 of *Lecture Notes in Computer Science*, pages 31–42. Springer Berlin Heidelberg, 2010.

[Bel64] John Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195–200, 1964.

[Bel66] John Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447, 1966.

[Bel87] John Stewart Bell. *Speakable and unspeakable in quantum mechanics*. Collected papers on quantum philosophy. Cambridge Univ. Press, Cambridge, 1987.

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

[BGSS07] Nicolas Brunner, Nicolas Gisin, Valerio Scarani, and Christoph Simon. Detection loophole in asymmetric bell experiments. *Physical review letters*, 98(22):220403, 2007.

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

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

[BLM+05] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, 2005.

[BM03] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

[Boh35] Niels Bohr. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 48:696–702, 1935.

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

[Bor99] Brian Borchers. Csdp, ac library for semidefinite programming. *Optimization methods and Software*, 11(1-4):613–623, 1999.

[BP05] Jonathan Barrett and Stefano Pironio. Popescu-rohrlich correlations as a unit of nonlocality. *Physical Review Letters*, 95(14):140401, 2005.

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

[BRG+13] Fernando GSL Brandão, Ravishankar Ramanathan, Andrzej Grudka, Karol Horodecki, MichaL Horodecki, and PaweL Horodecki. Robust device-independent randomness amplification with few devices. *arXiv preprint arXiv:1310.4544*, 2013.

[BW92] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884, November 1992.

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

[CBZG07] Ping-Xing Chen, János Bergou, Shi-Yao Zhu, and Guang-Can Guo. Ancilla dimensions needed to carry out positive-operator-valued measurement. *Phys. Rev. A*, 76:060303, Dec 2007.

[CDP11] Giulio Chiribella, Giacomo M. D'Ariano, and Paolo Perinotti. Informational derivation of quantum theory. *Physical Review A*, 84(1):012311+, July 2011.

[CG04] Daniel Collins and Nicolas Gisin. A relevant two qubit bell inequality inequivalent to the chsh inequality. *Journal of Physics A: Mathematical and General*, 37(5):1775, 2004.

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

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

[Chu36] Alonzo Church. An unsolvable problem of elementary number theory. *The American Journal Of Mathematics*, 58:345–363, 1936.

[CK06] John Conway and Simon Kochen. The free will theorem. *Foundations of Physics*, 36(10):1441–1473, 2006.

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

[CR11] Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nat Commun*, 2:411–, August 2011.

[CR12a] R. Colbeck and R. Renner. The completeness of quantum theory for predicting measurement outcomes. *arXiv*, arXiv:1208.4123 [quant-ph], 2012.

[CR12b] Roger Colbeck and Renato Renner. Free randomness can be amplified. *Nat Phys*, 8(6):450–454, June 2012.

[CW12] EricG. Cavalcanti and HowardM. Wiseman. Bell nonlocality, signal locality and unpredictability (or what bohr could have told einstein at solvay had he known about bell experiments). 42(10):1329–1338–, 2012.

[DB11] B. Dakic and C. Brukner. Quantum theory and beyond: Is entanglement special? pages 365–392, 2011.

[DdlTA14] Chirag Dhara, Gonzalo de la Torre, and Antonio Acín. Can observed randomness be certified to be fully intrinsic? *Physical review letters*, 112(10):100402, 2014.

[DH10] Rod Downey and Denis Hirschfeldt. *Algorithmic Randomness and Complexity*. Theory and Applications of Computability Series. Springer, 2010.

[dlTHD+14] Gonzalo de la Torre, Matty J Hoban, Chirag Dhara, Giuseppe Prettico, and Antonio Acín. Maximally nonlocal theories cannot be maximally random. *arXiv preprint arXiv:1403.3357*, 2014.

[dlTMSM12] Gonzalo de la Torre, Llu'is Masanes, Anthony J. Short, and Markus P. Müller. Deriving Quantum Theory from Its Local Structure and Reversibility. *Physical Review Letters*, 109:090403+, August 2012.

[DM03] Jonathan P. Dowling and Gerard J. Milburn. Quantum technology: the second quantum revolution. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1809):1655–1674, August 2003.

[DPA13] Chirag Dhara, Giuseppe Prettico, and Antonio Acín. Maximal quantum randomness in bell tests. *Physical Review A*, 88(5):052116, 2013.

[DPP05] Giacomo Mauro D'Ariano, Paoloplacido Lo Presti, and Paolo Perinotti. Classical randomness in quantum measurements. *Journal of Physics A: Mathematical and General*, 38(26):5979, 2005.

[DPS05] Andrew Doherty, Pablo Parrilo, and Federico Spedalieri. Detecting multipartite entanglement. *Phys. Rev. A*, 71:032333, Mar 2005.

[EHGC04] Jens Eisert, Philipp Hyllus, Otfried Gühne, and Marcos Curty. Complete hierarchies of efficient approximations to problems in entanglement theory. *Phys. Rev. A*, 70:062317, Dec 2004.

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

[FFW11] T. Franz, F. Furrer, and R. F. Werner. Extremal quantum correlations and cryptographic security. *Phys. Rev. Lett.*, 106:250502, 2011.

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

[FN13] Santiago Figueira and André Nies. Feasible analysis, randomness and base invariance. *Theory of Computing Systems*, 2013. To appear.

[Fri12] Tobias Fritz. Tsirelson's problem and kirchberg's conjecture. *Reviews in Mathematical Physics*, 24(05), 2012.

[Fro81] M. Froissart. Constructive generalization of bell's inequalities. *Il Nuovo Cimento B*, 64(2):241–251, 1981.

[FS14] Santiago Figueira and Gabriel Senno. Four lectures on algorithmic randomness. *Internal notes of a series of talks delivered at ICFO*, 2014.

[FSA+13] Tobias Fritz, Ana Belén Sainz, Remigiusz Augusiak, J Bohr Brask, Rafael Chaves, Anthony Leverrier, and A Acín. Local orthogonality as a multipartite principle for quantum correlations. *Nature communications*, 4, 2013.

[Gha10] Sevag Gharibian. Strong np-hardness of the quantum separability problem. *Quantum Info. Comput.*, 10(3):343–360, March 2010.

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

[Gis07] Nicolas Gisin. Bell inequalities: many questions, a few answers. *arXiv preprint quant-ph/0702021*, 2007.

[GMCD10] David Gross, Markus Mueller, Roger Colbeck, and Oscar C. O. Dahlsten. All reversible dynamics in maximally non-local theories are trivial. March 2010.

[GMDLT+13] Rodrigo Gallego, Lluis Masanes, Gonzalo De La Torre, Chirag Dhara, Leandro Aolita, and Antonio Acin. Full randomness from arbitrarily deterministic events. *Nature communications*, 4, 2013.

[Gol67] E Mark Gold. Language identification in the limit. *Information and control*, 10(5):447–474, 1967.

[Gur04] Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448 – 484, 2004. Special Issue on {STOC} 2003.

[GWAN12] Rodrigo Gallego, Lars Erik Würflinger, Antonio Acín, and Miguel Navascues. Operational framework for nonlocality. *Phys. Rev. Lett.*, 109(7):070401–, August 2012.

[Haa92] R. Haag. *Local quantum physics: fields, particles, algebras*. Texts and Monographs in Physics. Springer-Verlag, 1992.

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

[Hal11] Michael J. W. Hall. Relaxed bell inequalities and kochen-specker theorems. *Phys. Rev. A*, 84(2):022102–, August 2011.

[Har01] Lucien Hardy. Quantum Theory From Five Reasonable Axioms, September 2001.

[Har09a] Lucien Hardy. Foliable Operational Structures for General Probabilistic Theories, December 2009.

[Har09b] Aram Wettroth Harrow. Exact universality from any entangling gate without inverses. *Quantum Information Computation*, 9(9):773–777, 2009.

[Har13] Lucien Hardy. Reconstructing quantum theory, March 2013.

[HCLB11] M. J. Hoban, E. Campbell, K. Loukopoulos, and D. Browne. Non-adaptive measurement-based quantum computation and multi-party bell inequalities. *New J. Phys.*, 13:023014, 2011.

[HdV13] Marcus Huber and Julio de Vicente. Structure of multidimensional entanglement in multipartite systems. *Phys. Rev. Lett.*, 110:030501, Jan 2013.

[HHH96] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223(1):1 – 8, 1996.

[HJ12] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[HM04] J Helton and Scott McCullough. A positivstellensatz for noncommutative polynomials. *Transactions of the American Mathematical Society*, 356(9):3721–3737, 2004.

[HW10] Lucien Hardy and William K. Wootters. Limited Holism and Real-Vector-Space Quantum Theory, May 2010.

[JGBB11] Peter Janotta, Christian Gogolin, Jonathan Barrett, and Nicolas Brunner. Limits on nonlocal correlations from the structure of the local state space. *New Journal of Physics*, 13(6):063024, 2011.

[JM05] Nick S. Jones and Lluis Masanes. Interconversion of nonlocal correlations. *Phys. Rev. A*, 72:052312, 2005.

[JNP+11] Marius Junge, Miguel Navascues, Carlos Palazuelos, D Perez-Garcia, Volkher B Scholz, and Reinhard F Werner. Connes' embedding problem and tsirelson's problem. *Journal of Mathematical Physics*, 52(1):012102, 2011.

[KHS+12] D. E. Koh, M. J. W. Hall, Setiawan, J. E. Pope, C. Marletto, A. Kay, V. Scarani, and A. Ekert. The effects of reduced "free will" on Bell-based randomness expansion, 2012.

[Kle67] Stephen Cole Kleene. *Mathematical Logic*, volume 18. Dover Publications, 1967.

[KPB06] Johannes Kofler, Tomasz Paterek, and ÄŒaslav Brukner. Experimenter'ss freedom in bell's theorem and quantum cryptography. *Phys. Rev. A*, 73(2):022104–, February 2006.

[Kup] Greg Kuperberg. Complexity zoo active inclusion diagram. `https://www.math.ucdavis.edu/~greg/zoology/diagram.xml`.

[Lan92] R. Landauer. Information is physical. In *Physics and Computation, 1992. PhysComp '92., Workshop on*, pages 1 –4, oct 1992.

[Lap40] P. S. Laplace. *A philosophical essay on probabilities.* 1840.

[Las01] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[Lev73] Leonid A Levin. On the notion of a random sequence. *Soviet Math. Dokl*, 14(5):1413–1416, 1973.

[LKPR10] Jonathan Lavoie, Rainer Kaltenbaek, Marco Piani, and Kevin J. Resch. Experimental bound entanglement in a four-photon state. *Phys. Rev. Lett.*, 105:130501, Sep 2010.

[Lof04] Johan Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289. IEEE, 2004.

[LPY+12] Hong-Wei Li, Marcin Pawlowski, Zhen-Qiang Yin, Guang-Can Guo, and Zheng-Fu Han. Semi-device-independent randomness certification using n $\rightarrow$ 1 quantum random access codes. *Phys. Rev. A*, 85:052308, May 2012.

[LV08] Ming Li and Paul Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer Verlag, 2008.

[LV09] Ming Li and Paul MB Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer Science & Business Media, 2009.

[LVB11] Yeong-Cherng Liang, Tamás Vértesi, and Nicolas Brunner. Semi-device-independent bounds on entanglement. *Phys. Rev. A*, 83(2):022108–, February 2011.

[Mac63] G.W. Mackey. *The mathematical foundations of quantum mechanics: a lecture-note volume*. Mathematical physics monograph series. W.A. Benjamin, 1963.

[MAG06] Ll. Masanes, A. Acin, and N. Gisin. General properties of nonsignaling theories. *Phys. Rev. A*, 73(1):012112–, January 2006.

[Mar66] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602–619, 1966.

[Mas05] Ll Masanes. Extremal quantum correlations for n parties with two dichotomic observables per site. *arXiv preprint quant-ph/0512100*, 2005.

[Mas09] Lluis Masanes. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.*, 102(14):140501–, April 2009.

[MBL+13] Tobias Moroder, Jean-Daniel Bancal, Yeong-Cherng Liang, Martin Hofmann, and Otfried Gühne. Device-independent entanglement quantification and related applications. *Phys. Rev. Lett.*, 111:030501, Jul 2013.

[Mer90] N. David Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.*, 65(15):1838–1840, October 1990.

[MM11] Lluis Masanes and Markus P. Mueller. A derivation of quantum theory from physical requirements. *New Journal of Physics*, 13(6):063001+, May 2011.

[MMPGA14] Ll Masanes, Markus P Mueller, David Perez-Garcia, and Remigiusz Augusiak. Entanglement and the three-dimensionality of the bloch ball. *Journal of Mathematical Physics*, 55(12):122203, 2014.

[MOD12] Markus P. Mueller, Jonathan Oppenheim, and Oscar C. O. Dahlsten. The black hole information problem beyond quantum theory. June 2012.

[MPA11] Lluis Masanes, Stefano Pironio, and Antonio Acin. Secure device-independent quantum key distribution with causally independent measurement devices. *Nat Commun*, 2:238–, March 2011.

[MS13] Carl A. Miller and Yaoyun Shi. Optimal Robust Self-Testing by Binary Nonlocal XOR Games. In Simone Severini and Fernando Brandao, editors, *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)*, volume 22 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 254–262, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000.

[NdlTV14] Miguel Navascués, Gonzalo de la Torre, and Tamás Vértesi. Characterization of quantum correlations with local dimension constraints and its device-independent applications. *Physical Review X*, 4(1):011011, 2014.

[NGHA14] Miguel Navascués, Yelena Guryanova, Matty J Hoban, and Antonio Acín. Almost quantum correlations. *arXiv preprint arXiv:1403.4621*, 2014.

[NOP09] Miguel Navascués, Masaki Owari, and Martin B Plenio. Power of symmetric extensions for entanglement detection. *Physical Review A*, 80(5):052306, 2009.

[NPA07] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98(1):010401–, January 2007.

[NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013–, 2008.

[NV14] Miguel Navascues and Tamas Vertesi. Bounding the set of finite dimensional quantum correlations. *arXiv preprint arXiv:1412.0924*, 2014.

[Odi99] P. Odifreddi. *Classical recursion theory*, volume 2 of *Studies in logic and the foundations of mathematics*. Elsevier, 1999.

[PAB+09] Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009.

[PAM+10a] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by bell's theorem. *Nature*, 464:1021, 2010.

[PAM+10b] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by bell's theorems. *Nature*, 464(7291):1021–1024, April 2010.

[PAM+10c] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzimitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al. Random numbers certified by Bell's theorem. *Nature*, 464(7291):1021–1024, 2010.

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[PB11] Marcin PawLowski and Nicolas Brunner. Semi-device-independent security of one-way quantum key distribution. *Phys. Rev. A*, 84:010302, Jul 2011.

[PBS11] Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. Extremal correlations of the tripartite no-signaling polytope. *Journal of Physics A: Mathematical and Theoretical*, 44(6):065303–, 2011.

[PCL+12] Jian W. Pan, Zeng B. Chen, Chao Y. Lu, Harald Weinfurter, Anton Zeilinger, and Marek Żukowski. Multiphoton entangle-

ment and interferometry. *Reviews of Modern Physics*, 84:777–838, May 2012.

[Per96] Asher Peres. Collective tests for quantum nonlocality. *Phys. Rev. A*, 54(4):2685–2689, October 1996.

[Per99] Asher Peres. All the bell inequalities. *Foundations of Physics*, 29(4):589–614, April 1999.

[PGWP+08] D. Pérez-García, M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge. Unbounded violation of tripartite bell inequalities, 2008.

[PM11] S. Pironio and S. Massar. Security of practical private randomness generation, 2011.

[PM13] Stefano Pironio and Serge Massar. Security of practical private randomness generation. *Phys. Rev. A*, 87:012336, Jan 2013.

[PMLA13] S. Pironio, Ll. Masanes, A. Leverrier, and A. Acin. Security of device-independent quantum key distribution in the bounded-quantum-storage model. *Phys. Rev. X*, 3:031007, Aug 2013.

[PPK+09] Marcin Pawlowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Zukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, October 2009.

[PR94] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, March 1994.

[Pus13] Matthew Pusey. Negativity and steering: A stronger peres conjecture. *Phys. Rev. A*, 88:032313, Sep 2013.

[PV08] Károly F. Pál and Tamás Vértesi. Efficiency of higher-dimensional hilbert spaces for the violation of bell inequalities. *Phys. Rev. A*, 77(4):042105–, April 2008.

[PV09a] Károly Pál and Tamás Vértesi. Concavity of the set of quantum probabilities for any given dimension. *Phys. Rev. A*, 80:042114, Oct 2009.

[PV09b] Károly Pál and Tamás Vértesi. Quantum bounds on bell inequalities. *Phys. Rev. A*, 79:022120, Feb 2009.

[PV09c] Károly F. Pál and Tamás Vértesi. Concavity of the set of quantum probabilities for any given dimension. *Phys. Rev. A*, 80:042114, Oct 2009.

[PV10] Károly Pál and Tamás Vértesi. Maximal violation of a bipartite three-setting, two-outcome bell inequality using infinite-dimensional quantum systems. *Phys. Rev. A*, 82:022116, Aug 2010.

[Roc70] R.T. Rockafellar. *Convex Analysis*. Princeton mathematical series. Princeton University Press, 1970.

[RUV13] Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013.

[Sca14] Valerio Scarani. Private communication. 2014.

[Sch71] C.P. Schnorr. A unified approach to the definition of random sequences. *Mathematical systems theory*, 5(3):246–258, 1971.

[Sch72] C. P. Schnorr. The process complexity and effective random tests. In *Proceedings of the Fourth Annual ACM Symposium on Theory of Computing*, STOC '72, pages 168–176, New York, NY, USA, 1972. ACM.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[Sho94] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Statist. Comput.*, 26:1484, 1994.

[Soa87] Robert I. Soare. *Recursively enumerable sets and degrees*. Springer, Heidelberg, 1987.

[Sol64a] Ray J Solomonoff. A formal theory of inductive inference. Part I. *Information and control*, 7(1):1–22, 1964.

[Sol64b] Ray J Solomonoff. A formal theory of inductive inference. Part II. *Information and control*, 7(2):224–254, 1964.

[Stu99] Jos F Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *J. Comput. Syst. Sci.*, 33(1):75–87, 1986.

[SW08] Volkher B Scholz and Reinhard F Werner. Tsirelson's problem. *arXiv preprint arXiv:0812.4305*, 2008.

[Tsi80] Boris. S. Tsirelson. Quantum generalization of bell's inequality. *Lett. Math. Phys.*, 4:93, 1980.

[TSS13] Le Thinh, Lana Sheridan, and Valerio Scarani. Bell tests with min-entropy sources. *Phys. Rev. A*, 87:062121, Jun 2013.

[Tur36] Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 42:230–265, 1936.

[TW01] Max Tegmark and John A. Wheeler. 100 Years of the Quantum. January 2001.

[VB14] Tamás Vértesi and Nicolas Brunner. Disproving the peres conjecture by showing bell nonlocality from bound entanglement. *Nat Commun*, 5, 11 2014.

[VN11] Tamás Vértesi and Miguel Navascués. Certifying entangled measurements in known hilbert spaces. *Phys. Rev. A*, 83:062112, Jun 2011.

[VP08] T. Vértesi and K. F. Pál. Generalized clauser-horne-shimony-holt inequalities maximally violated by higher-dimensional systems. *Phys. Rev. A*, 77(4):042106–, April 2008.

[VPB10] Tamás Vértesi, Stefano Pironio, and Nicolas Brunner. Closing the detection loophole in bell experiments using qudits. *Physical review letters*, 104(6):060401, 2010.

[VV12a] U. Vazirani and T. Vidick. Certifiable quantum dice: or, true random number generation secure against quantum adversaries. *Proceedings of the ACM Symposium on the Theory of Computing*, 2012.

[VV12b] Umesh V. Vazirani and Thomas Vidick. Fully device independent quantum key distribution, 2012.

[VW02] G. Vidal and R. F. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65(3):032314–, February 2002.

[WCD08] Stephanie Wehner, Matthias Christandl, and Andrew C. Doherty. Lower bound on the dimension of a quantum system given measured data. *Phys. Rev. A*, 78(6):062112–, December 2008.

[WW01] R. F. Werner and M. M. Wolf. Bell inequalities and entanglement. *Quantum Information and Computation*, 1:1–25, Oct 2001.

[ZST10] Xin-Yuan Zhao, Defeng Sun, and Kim-Chuan Toh. A newton-cg augmented lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.

[ZZ08] Thomas Zeugmann and Sandra Zilles. Learning recursive functions: A survey. *Theoretical Computer Science*, 397(1):4–56, 2008.