Advances in quantum key distribution and quantum randomness generation

LE PHUC THINH
(B.Sc.(Hons.), NUS)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosopy

in the

Centre for Quantum Technologies
National University of Singapore

2015

DECLARATION

I hereby declare that this thesis is my original work and has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Le Phuc Thinh

April 7, 2015

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor Valerio Scarani for his expert guidance, without which the work in this thesis could not have been possible, and his friendship since my undergraduate years. His deep intuitions, insights and approach to scientific research have greatly influenced my research.

Secondly, I would like to thank all my friends and collaborators who have made my life more chaotic both quantumly and classically. My sincere thank to Lana Sheridan, Jean Daniel-Bancal, Eduardo Martin-Martinez, Marco Tomamichel and Stephanie Wehner for teaching me so much throughout the years, Charles Lim for being a good friend who made my trajectory and quantum information intersect, and Le Huy Nguyen, Cai Yu, Rafael Rabelo, Melvyn Ho for sharing the office with me and making my daily life more fun. Thank you Yang Tzyh Haur, Colin Teo, Haw Jing Yan, Jiri Minar, Wang Yimin, Wu Xing Yao, Alexandre Roulet, Law Yun Zhi, Goh Koon Tong, Nelly Ng, Jedrzej Kaniewski for sharing memories and helps, and the CQT staffs for providing the perfect research environment. Not forgetting Nicolas Gisin, Hugo Zbinden, Stefano Pironio, Nicolas Brunner, Marcos Curty, Tobias Moroder and Gonzalo de la Torre for the stimulating discussions and hospitality.

And to all who has helped me in one way or another, let it be known that I will always remember and cherish your help and friendship.

I thank my PhD examiners Thomas Vidick, Roger Colbeck and Dagomir Kaszlikowski for their helpful comments on an earlier version of this thesis.

Finally, I would like to specially thank my parents for their continuous support and education, and without whom my entire timeline would have never existed. This thesis is fully dedicated to my parents.

ABSTRACT

Quantum information science has completely changed the way we think about and process information. From the simple realization that *information is physical*, we have been able to use the peculiar features of quantum mechanical phenomena to our advantage. Designing algorithms whose performance exceeds those running on classical computers, and performing secret communication whose security can actually be proven from sound assumptions are two main catalysts for the establishment of the field.

This thesis discusses some progress in quantum key distribution and quantum randomness generation. Quantum key distribution, which is motivated by the increasing need to communicate securely, is on the verge of becoming an established technology. With the goal of extending the reach of quantum key distribution to more realistic scenarios, we present a study on reference-frame-independent protocols whose knowledge can help design more efficient protocols, and a framework to the security analysis of distributed-phase-reference protocols, which have been missing for many years. This allows these protocols to be used in practice against the most general adversary, although the key rate is rather pessimistic. In quantum randomness generation, the amount of extractable randomness from a quantum system depends on the level of trust or characterization of the devices; we present a study into such interaction. In the extreme situation of distrust, i.e. device-independent scenarios, we study the effect of the input randomness on the certifying power of such scenarios, and realize that one cannot amplify an arbitrary min-entropy source device-independently. Finally we discuss the amount of randomness in post-selected data, which has consequences on practical randomness generation protocols. The in depth study of randomness generation from quantum processes is well justified by the important role of randomness in modern computer science and other fields.

CONTENTS

CHAPTER 1

INTRODUCTION

Since its birth in 1920's, quantum mechanics has been very successful at predicting and explaining phenomena happening in the microscopic world. Despite its tremendous success, deep philosophical and conceptual questions related to the foundation of quantum mechanics linger to the present day [1]. However, as researchers wrestle with these difficulties a paradigm shift slowly happens: it is realized that the mind-boggling quantum weirdness can actually have practical applications in computer science and engineering.

The first example is quantum cryptography, or more precisely quantum key distribution [2]. First proposed by Charles H. Bennett and Gilles Brassard in 1984 [3] and later by Artur Ekert in 1990, quantum key distribution offers a solution to the key distribution problem in classical private key cryptosystems such as the one-time-pad. The solution is an ingenious spin on the standard "problems" with quantum mechanics, utilizing these negativities to our advantage. Because one cannot measure without disturbing and cannot duplicate an unknown quantum system, they serve as ideal information couriers to carry the key between distant parties. Any attempt at eavesdropping ultimately manifests as errors which the parties can detect; therefore the security of the key is guaranteed by principles of quantum mechanics.

The second example is quantum randomness. In contrast to classical mechanics, being probabilistic is the norm in quantum mechanics. This feature left Einstein wonder if there may exist hidden variables such that when discovered would explain the probabilistic nature of quantum mechanics and return us to the deterministic worldview [4]. The apparently philosophical issue is conclusively answered by John Bell in his discovery of Bell inequalities [5]. When a Bell inequality is violated in an experiment as demonstrated in [6, 7], the results are intrinsically random: no

local hidden variables can explain the results of such experiments. In other words, quantum mechanics can be used to generate randomness and we have again utilize the strange features of quantum mechanics to our advantage! Incidentally, the power of Bell does not stop there; it propels the field of quantum non-locality and the device-independent approach into existence [8].

It is these developments that open up a new interdisciplinary field of scientific investigation known today as quantum information science, which comprises of many subfields notably quantum computing, quantum communication, quantum information theory, and the aforementioned quantum cryptography. This thesis presents some recent theoretical advancements in quantum key distribution and quantum randomness generation, the motivations for which we briefly discuss next.

Quantum cryptography is born out of the need to communicate secretly. While secure communication is an obvious need of governments and corporations, the daily consumers of internet services are not entirely safe from spying eyes, in light of increasing instances of hacking and surveillance. Therefore, in order to communicate securely, one must employ techniques of cryptography, the science and art of rendering a message unintelligible to any unauthorized party [9]. Cryptographic systems, or methods for encryption and decryption of messages, fall into two main categories: public and private key cryptosystems. The security of public key cryptosystems such as RSA [10] relies on the computational complexity of prime factorization, whereas that of private key cryptosystems such as one-time-pad only rests on the security of a common secret key. Security based on factorization complexity is unlikely to withstand challenges posed by the development of fast quantum computers in the future [11]. On the contrary, it is proven that the one-time-pad cryptosystem is information-theoretically secured provided the key is truly random, as long as the message, used only once and unknown to any unauthorized party [12]. Hence, one-time-pad cryptosystem provides an ideal method for secret communication if the problem of key distribution is solved. An obvious solution to the key distribution problem is for the two communicating parties to meet and agree upon a secret key. However, it is clear that their secret communication can only be sustained until they use up their pre-established key. They may think of using a trusted courier to deliver the key but finding such a trustworthy agent is certainly not an easy task because classical agents are prone to corruption. Moreover, they have to tackle the problem of key storage before the encryption when sending a message, especially when the key is very long and must be kept secret for extended period of time. Quantum key distribution offers a nice solution to the key distribution problem. With the use of quantum mechanical systems as information carriers we can guarantee the security of the secret key

based on our understanding of the law of quantum physics. Furthermore, the key can be distributed on demand before secret communication is required, which eliminates partly the problem of key storage before secret communication. The best known example of a QKD protocol is the BB84 protocol proposed by Charles H. Bennett and Gilles Brassard in 1984.

Randomness is an important concept and also a fundamental resource in modern science. It is used to assign test subjects in a randomized controlled trials so scientists can test their hypothesis, or to randomly select a sample out of a population for analysis to avoid experimental design bias. It is present in the analysis of experiments, e.g. to see if a certain effect is due to chance or has an underlying cause, and used in randomized algorithms and statistical simulations, etc. It lies at the heart of cryptography, and a close analysis of quantum key distribution protocols we have just introduced reveals that it is used there as well. Randomness is also essential to the operation of casinos. It is thus crucial to investigate methods of generating randomness. However, the notions of randomness used in applications are not equal; it can roughly be categorized based on whether the randomness is required to be private, as in cryptographic and gambling applications or not, as in the other remaining applications. In other words, randomness can appear to be perfect but when used in such applications, renders cryptographic insecurity or a loss to the casinos. Quantum mechanical processes therefore serve as the best known candidates to date to generate private randomness, and has been the subject of several experimental proposals. The task of private randomness generation is first explored in Roger Colbeck's thesis [13].

Since their conceptions, both fields have undergone significant development. The main problem in the beginning of quantum key distribution was to obtain a rigorous security analysis of the BB84 protocol and its variants such as the six-state protocol [14, 15]; some proofs were rather technical [16, 17] while some required the link between privacy amplification, entanglement purification and quantum error correction [18, 19]. Later, by noticing that most quantum cryptography protocols (BB84 included) are permutation invariant, the analysis of a general protocol was simplified tremendously. One only need to consider security against a much more restricted class of attacks known as the collective attacks where Eve interacts with each quantum signals using the same strategy [20, 21]. At the same time, it was realized that the security definition used in several works were not satisfactory, i.e. not composable and may undermine the security of an application where quantum key distribution is used as a subprotocol [22]. Then advancements were made on the finite key security proof using ideas such as the quantum deFinetti theorem [23], post-selection technique [24] and the entropic uncertainty

relations [25]. Today, experimentalists and theorists are working closely together to bridge the gap between theoretical modeling and experimental realization. This effort has spawned further ideas such as measurement-device-independent protocol [26] or device-independent quantum key distribution [27]. For private randomness generation, after the first investigation by Colbeck, the field quickly developed along two main directions: randomness expansion and randomness amplification. In randomness expansion, the main goal is to expand a small amount of high quality randomness; one of the first paper to consider this task in the device-independent setting is [28] which holds for adversary holding classical information. Security against quantum adversary as well as better expansion (up to unbounded expansion) were developed later [29–31]. In randomness amplification, we start with low quality randomness and try to make it more uniform or more perfect. The amplification of Santha-Vazirani source using quantum resource was first studied in [32] but the result was limited to relatively high quality sources. Later [33] extended the result to arbitrary weak Santha-Vazirani source. Since then, further results on amplification against quantum adversaries and amplification of min-entropy sources were obtained; consult [34] for a review of these results.

We may wonder how can one contribute to such a developed field? Fortunately, even though a lot of progress has been made, there are still many open problems to consider. For instance, how can we perform quantum key distribution in practical scenarios such as earth-to-satellite (in anticipation of the development of a global quantum internet) and chip-to-chip communication while respecting and utilizing all the scenarios' constraints? Also despite major development and understanding of security proofs, the security proof for a certain class of quantum cryptography protocols called the distributed-phase-reference protocols is still missing because of the lack of permutation invariance. In the field of quantum randomness amplification, it is still unknown whether one can amplify weaker randomness source such as the min-entropy source. Moreover, an understanding of the payoff between different assumptions on the randomness generation scenario or the way we post-process the experimental outcomes and the amount of randomness obtained is still lacking. This thesis provides partial answers to such questions.

Contributions

In the field of quantum key distribution, we make two important advancements. Firstly, in recent years there has been an interest in the reference frame independent protocol proposed by Laing *et al.* [35], which enable the distribution of secret keys without the need to align the reference frame of the experimental setups of Alice

and Bob. While trying to generalize the protocol to d level quantum systems, we realized that the protocol is actually doing tomography in disguise and from such tomographic information one can have better key rate or even realign the reference frames than by passing through the parameter C. Secondly, the security proofs of distributed phase reference protocols have been restricted to single photon sources or specific attacks [36–38]. Using the deFinetti approach, we provide a security proof against the most general adversary, namely one who can perform arbitrary coherent attacks on the signals. Our security proof relies on numerical methods to bound the error rates from the observed data and may be of independent interest.

In the field of quantum randomness generation, there are two main tasks as mentioned before: randomness expansion and randomness amplification. Although many strong results have been obtained in the literature, the assumptions involved are often implicit in the proof. Here we provide an analysis on various conceivable scenarios, which helps clarify various concepts and provides an overall understanding of the task of randomness generation from quantum systems. Our framework leads to various bounds on the amount of randomness which depend on the assumptions made. We also consider the task from the point of view of practical experiments where photonic implementations suffer from a lot of no-detection events. Our contribution here involves obtaining a correct bound for the amount of randomness in the post-selected events consisting of the detected runs, which benefits the classical post-processing.

The task of randomness amplification (without the use of an independent seed) has received a lot of attention recently. It is well known that one cannot amplify a single Santha-Vazirani source or min-entropy source classically. However, it was first proven by Colbeck and Renner that one can amplify a Santha-Vazirani source of high enough quality using quantum resources. This direction has been completed by Gallego *et al.* in [33] where any arbitrary Santha-Vazirani source can be amplified with a five partite Bell scenario. Nevertheless, the question of amplifying min-entropy sources using quantum resource remains open. Here we prove a general impossibility result: it is not possible to amplify arbitrary min-entropy sources by using arbitrary no-signalling resources. Our result is compatible with other works in the literature; for instance the amplification protocol [39] assumes the initial min-entropy is relatively high.

Overview of the thesis

The rest of this thesis is divided into four chapters.

Chapter 2: Preliminaries

In this chapter we present the basic background material underlying the thesis. We first introduce some basic quantum information concepts and tools and then discuss some basic background on Bell nonlocality. This chapter also serves to establish some notation used in this thesis.

Chapter 3: Quantum Key Distribution

The chapter starts with a brief introduction to quantum cryptography and move on to discuss the security definitions and extractable key rate. Then we move on to present the results on reference-frame-independent protocols, which are based on the paper [40] and a framework to prove the security of distributed-phase-reference protocols against coherent attacks [41].

Chapter 4: Quantum Randomness Generation

We first lay out scenarios for quantum randomness generation which is based on the levels of characterization (or trust) of the devices. Then we present a study on the relationship between devices' levels of characterization and randomness generation [42], which assumes the measurement independent assumption. Dropping this assumption, we investigate the role of the input randomness in Bell tests. This result, which is based on the paper [43], has important consequences on both device-independent applications and as well as foundations of physics. Our final result is about the amount of randomness present in a subset of post-selected events [44]. The study is motivated by the need to discard the double no-detection events which occur very often in Bell tests because of the inefficiencies of the source and detectors.

Chapter 5: Conclusions and Outlook

This chapter concludes the thesis and gives several remarks on the possible future directions of the field.

CHAPTER 2

PRELIMINARIES

Most of the materials in this chapter are basic working knowledge in quantum information. The readers are referred to [45] for an introduction to quantum information science, [23] for an introduction to the tools used in quantum key distribution, and [8] for a recent review on Bell nonlocality.

2.1 Mathematical notations

Unless otherwise stated, all Hilbert spaces are assumed to be finite dimensional. The state of a quantum system is described by a positive semidefinite operator of trace one acting on some Hilbert space $\mathcal{H}$. For convenience, we may use subnormalized states, which are positive semidefinite operators having trace at most one; the set of states and subnormalized states are denoted as $\mathcal{S}_=(\mathcal{H})$ and $\mathcal{S}_\leq(\mathcal{H})$, respectively. If the state is diagonal in some basis $\rho = \sum_x \lambda_x |x\rangle \langle x|$, then the system is said to be classical in this basis and a simpler description by a probability distribution $P_X(x) = \lambda_x$ suffices. Conversely, classical system can be described in the quantum formalism as a state diagonal in some basis. For composite systems, the joint Hilbert space is given by the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ and given a joint state ρ_{AB} the reduced state of each subsystem is given by the partial trace operation $\rho_A = \mathrm{tr}_A(\rho_{AB})$. A classical-quantum state describes the correlation between a classical (in some basis) and a quantum system, and is of the form $\rho_{XE} = \sum_x P_X(x) |x\rangle \langle x| \otimes \sigma_E^x$, where the superscript in σ_E^x is used as a label to mean the quantum state of system E conditioned on the first system being x. Following this line of thought, a classical-classical system is described by a joint probability distribution $P_{XY}(x, y)$.

We will need two operator norms, the Schatten 1-norm and 2-norm. The 1-norm of an operator L is the sum of its singular values, namely $\|L\|_1 = \mathrm{tr}(|L|)$ where $|L| = \sqrt{L^\dagger L}$ is the unique positive square root of $L^\dagger L$. The 1-norm induces a metric or distance function known as the trace distance

$$D(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1 , \tag{2.1}$$

which reduces to the statistical distance or the total variational distance when both states are diagonal in the same basis

$$D\left(\sum_x P_X(x) |x\rangle \langle x| , \sum_x Q_X(x) |x\rangle \langle x|\right) = \frac{1}{2} \sum_x |P_X(x) - Q_X(x)|. \tag{2.2}$$

The 2-norm is induced by the Hilbert-Schmidt inner product $\langle A, B\rangle = \mathrm{tr}(A^\dagger B)$; explicitly, $\|L\|_2 = \sqrt{\langle L, L\rangle}$ and is the square root of the sum of the square of the singular values of L. Likewise, the 2-norm also induces a metric.

Entropies are measures of uncertainty. While there are many entropic quantities such as the α-Renyi entropies, the correct entropy which captures the worst-case uncertainty of an adversary Eve on some classical system (e.g. the key) is given by the conditional min-entropy. For quantum-quantum states, the conditional min-entropy is defined as

$$H_{\min}(A|B)_\rho = \max_\sigma \sup\{\lambda \in \mathbb{R} : \rho_{AB} \leq 2^{-\lambda} \mathbb{1}_A \otimes \sigma_B\} \tag{2.3}$$

where the condition $\rho_{AB} \leq 2^{-\lambda} \mathbb{1}_A \otimes \sigma_B$ means $2^{-\lambda} \mathbb{1}_A \otimes \sigma_B - \rho_{AB}$ is positive semidefinite and the maximization is taken over subnormalized quantum states. This quantity is related to the maximum fidelity to a maximally entangled state between A and B (which describes an omniscient observer B of A) one can recover by acting on half of the bipartite system [46]. In other words, the conditional min-entropy of ρ_{AB} is directly related to the maximum achievable singlet fraction [47]. For classical-quantum states ρ_{XE}, the min-entropy reduces to $-\log_2 P_{guess}(X|B)_\rho$ with the usual guessing probability of X given the quantum side information E

$$P_{guess}(X|E)_\rho = \max_{\mathcal{E}} \sum_x P_X(x) \langle x| \mathcal{E}(\sigma_E^x) |x\rangle \tag{2.4}$$

where the maximization is done over all quantum operation $\mathcal{E}$ (trace preserving completely positive maps). The smooth min-entropy is defined as a maximization of min-entropy over an ϵ-ball of states. Formally,

$$H_{\min}^\epsilon(A|B)_\rho = \max_{\tilde{\rho}} H_{\min}(A|B)_{\tilde{\rho}} \tag{2.5}$$

where $\tilde{\rho} \in \mathcal{P}(\rho, \epsilon) := \{\tau \in \mathcal{S}_{\leq}(\mathcal{H}_{AB}) : P(\rho, \tau) \leq \epsilon\}$ is the ϵ-ball around ρ with respect to the purified distance (a metric based on the fidelity). The smoothed entropies are the main quantity of interest in finite-size quantum information where statistical fluctuations only allow the estimation of ρ_{AB} up to some ϵ accuracy.

Dually, we have the notions of max-entropy and smooth max-entropy. They are defined as follows:

$$H_{\max}(A|C)_\rho = \min_\sigma \inf\{\lambda \in \mathbb{R} : \rho_{AC} \leq 2^\lambda \mathbb{1}_A \otimes \sigma_C\}, \tag{2.6}$$

$$H^\epsilon_{\max}(A|C)_\rho = \min_{\tilde{\rho}} H_{\max}(A|C)_{\tilde{\rho}}. \tag{2.7}$$

The max-entropy of quantum-quantum states is related to the decoupling accuracy, a quantity which captures how close a state ρ_{AC} to being "decoupled", namely being of the form $\mathbb{1}_A \otimes \sigma_C$ for arbitrary σ_C (describing an ignorant observer C of A). For classical-quantum states, the max-entropy is related to the security of a secret key.

The smooth min-entropy is important in randomness extraction and privacy amplification because of the following result [23]:

Theorem 1. ***(Leftover hash lemma against quantum side information)*** *Let ρ_{XE} be a classical-quantum state and $\mathcal{F}$ be a two-universal family of hash functions from $\mathcal{X}$ to $\mathcal{K} = \{0,1\}^\ell$. For any $0 \leq \epsilon \leq 1$,*

$$\frac{1}{2} \left\| \rho_{KEF} - U_K \otimes \rho_E \otimes U_F \right\|_1 \leq \epsilon + \frac{1}{2} \sqrt{2^{\ell - H^\epsilon_{\min}(X|E)}}, \tag{2.8}$$

where

$$\rho_{KEF} = \sum_f P_F(f) \rho_{f(X)E} \otimes |f\rangle\langle f| \tag{2.9}$$

is the final state after the action of a random $f \in \mathcal{F}$ chosen with uniform probability $P_F(f) = 1/|\mathcal{F}|$ and U_K is the completely mixed state of the system K.

Operationally, this result implies one can extract $H^\epsilon_{\min}(X|E)$ uniformly random bits from a source X, which may be correlated with an adversary E, by applying a randomly chosen function f from the two-universal family of hash functions to the output x of the process X.

2.2 Bell nonlocality

Bell's 1964 theorem represents one of the most profound developments in foundation of physics. In a typical Bell experiment, one finds two separated parties, each

performing measurements on their own system. The two systems may have interacted in the past: for instance they may be two photons emitted from the same source towards the two distant observers. After performing the measurements x and y chosen from a set of possibilities, they record the outcomes a and b, which may differ between runs even if the same measurements have been chosen. The data is used to interpret a family of conditional probability distributions $p(a, b|x, y)$ indexed by the pair of settings, which represents the "average behavior" of the experiment. Not surprisingly, there may be a correlation between the inputs and the outcomes and because of this reason, the family $p(a, b|x, y)$ is often called the correlation of the experiment. This correlation between distant parties certainly "cries out for explanation". A moment's thought lead Bell to a very plausible model which could explain the correlation: the local realistic model,

$$p(a, b|x, y) = \int_\Lambda \mathrm{d}\lambda p(\lambda) p(a|x, \lambda) p(b|y, \lambda), \tag{2.10}$$

where it is imagined, because two systems may have interacted in the past, that their behaviors are locally determined by a common hidden variable λ. Such behaviors form the local set $\mathcal{L}$ of correlations, i.e. a convex polytope for which tight Bell inequalities, for instance the CHSH inequality

$$|\langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \langle A_1 B_1\rangle| \leq 2 \tag{2.11}$$

with $a, b, x, y \in \{+1, -1\}$ and

$$\langle A_0 B_0\rangle = p(1, 1|0, 0) + p(-1, -1|0, 0) - p(1, -1|0, 0) - p(-1, 1|0, 0) \tag{2.12}$$

and similarly for other averages, are facets of the polytope. Correlations outside the local sets are the quantum correlations $\mathcal{Q}$, namely those which admit a representation

$$p(a, b|x, y) = \mathrm{tr}\left(\rho_{AB} M_{a|x} \otimes M_{b|y}\right) \tag{2.13}$$

for some state and measurements POVMs, and the no-signaling correlations $\mathcal{NS}$, i.e. $p(a, b|x, y)$ satisfying

$$\sum_b p(a, b|x, y) = \sum_b p(a, b|x, y') \qquad \text{for all } a, x, y, y' \tag{2.14}$$

$$\sum_a p(a, b|x, y) = \sum_a p(a, b|x', y) \qquad \text{for all } b, y, x, x' \tag{2.15}$$

which aim to capture the idea that any spacelike separated correlations cannot be used to signal a message between the two parties (for compatibility with Einstein's

theory of relativity).

A useful tool in any application of non-locality is the NPA hierarchy of semidefinite programs characterizing the quantum set of correlations [48, 49]. It is based on the following observation: first notice that in the defining condition for a quantum behavior (2.13), the quantum state and measurement can be "purified", namely we can assume the state to be pure and the measurements to be von Neumann projections. Let $\mathcal{O}$ be a set of k operators, define $\Gamma = (\Gamma_{ij})$ to be a $k \times k$ moment matrix (associated with $\mathcal{O}$) with entries $\Gamma_{ij} = \langle\psi| O_i^\dagger O_j |\psi\rangle$ for $O_i, O_j \in \mathcal{O}$ then it is clear that Γ is positive semidefinite: for all $\vec{u}$,

$$\vec{u}^\dagger \Gamma \vec{u} = \sum_{ij} u_i^* \Gamma_{ij} u_j = \langle\psi| \left(\sum_i u_i^* O_i^\dagger\right) \left(\sum_j u_j O_j\right) |\psi\rangle = \left\| \sum_j u_j O_j |\psi\rangle \right\|^2 \geq 0. \tag{2.16}$$

The observation applies to the choice of $\mathcal{O}$ being all the von Neumann projection operators $M_{a|x}, M_{b|y}$ together with the identity operator and some finite products of them. Moreover, the correlation $p(ab|xy)$ corresponds to a subset of the entries of Γ. Thus we have shown that if $p(ab|xy)$ is a quantum behavior then there exists a positive semidefinite moment matrix Γ which contains $p(ab|xy)$ as some entries. By restricting the maximum number n of operators in the allowed products we have an increasing sequence $\mathcal{O}_n$ (with respect to set inclusion) corresponding to a decreasing sequence of sets $\mathcal{Q}_n$ which better approximates the quantum set $\mathcal{Q}$.

As an example of the technique, let us describe the so called "local level 1" relaxation of the quantum set. In this case, the set $\mathcal{O}$ consists of $\mathbb{1}$, $A_{a|x}$, $B_{b|y}$, $A_{b|x}B_{b|y}$ for all possible choice a, b, x, y. The operators are not independent since for each measurement setting, say x for Alice, it must be that $\sum_a A_{a|x} = \mathbb{1}_A$; this allows us to eliminate dependent operators and obtain a simplified set $\mathcal{O}$. In particular, for the Bell scenario involving two parties each having two measurements with two outcomes, the associated operator set is

$$\mathcal{O} = \{\mathbb{1}, A_{0|0}, A_{0|1}, B_{0|0}, A_{0|0}B_{0|0}, A_{0|1}B_{0|0}, B_{0|1}, A_{0|0}B_{0|1}, A_{0|1}B_{0|1}\} \tag{2.17}$$

and therefore the (symmetric) moment matrix Γ is given by .

$$\begin{pmatrix}
1 & p_A(0|0) & p_A(0|1) & p_B(0|0) & p(00|00) & p(00|10) & p_B(0|1) & p(00|01) & p(00|11) \\
 & p_A(0|0) & v_1 & p(00|00) & p(00|00) & v_2 & p(00|01) & p(00|01) & v_3 \\
 & & p_A(0|1) & p(00|10) & v_2 & p(00|10) & p(00|11) & v_3 & p(00|11) \\
 & & & p_B(0|0) & p(00|00) & p(00|10) & v_4 & v_5 & v_6 \\
 & & & & p(00|00) & v_2 & v_5 & v_5 & v_7 \\
 & & & & & p(00|10) & v_6 & v_8 & v_6 \\
 & & & & & & p_B(0|1) & p(00|01) & p(00|11) \\
 & & & & & & & p(00|01) & v_3 \\
 & & & & & & & & p(00|11)
\end{pmatrix}$$

with v_j the unknown variables. If $p(ab|xy)$ is quantum then there must exist v_j such that $\Gamma \geq 0$, so this can be used to constraint the adversary to be in quantum, i.e. use quantum resources.

CHAPTER 3

QUANTUM KEY DISTRIBUTION

3.1 Introduction to QKD

3.1.1 The BB84 protocol

A QKD protocol is a set of instructions for the two distant parties, Alice and Bob, to generate a common secret key in the presence of an adversary Eve who is trying to learn about their key. The BB84 protocol uses four photon polarization states $|H\rangle$, $|V\rangle$, $|+45\rangle$, and $|-45\rangle$ belonging to the $+$ and $\times$ bases to encode and transmit information between Alice and Bob; $|H\rangle$ and $|+45\rangle$ codes for bit "0", while $|V\rangle$ and $|-45\rangle$ codes for bit "1". The protocol has 4 steps:

- Alice randomly prepares a photon in one of the four states and sends to Bob who will choose at random to measure it in either the $+$ or $\times$ basis. Each party will then have a list of pairs (bit, basis).
- Alice and Bob communicate over the classical channel the information of the basis of each bit, keeping only the bit which has been prepared and measured in the same basis (sifting).
- They announce a small subset of their bits in the $+$ and $\times$ bases to estimate the quantum bit error rate (QBER) or the probability of error in the bases (ϵ_z and ϵ_x, respectively).
- If the QBERs are not too large, they perform error correction to ensure that their list of bits are identical, otherwise they abort the protocol and no key is generated.

- They perform privacy amplification, reducing their list of bits to a shorter but more secured (unknown to any adversary Eve) common secret key. This is usually done by applying a random two-universal hash function, which can be chosen by Alice and communicated to Bob via the classical channel.

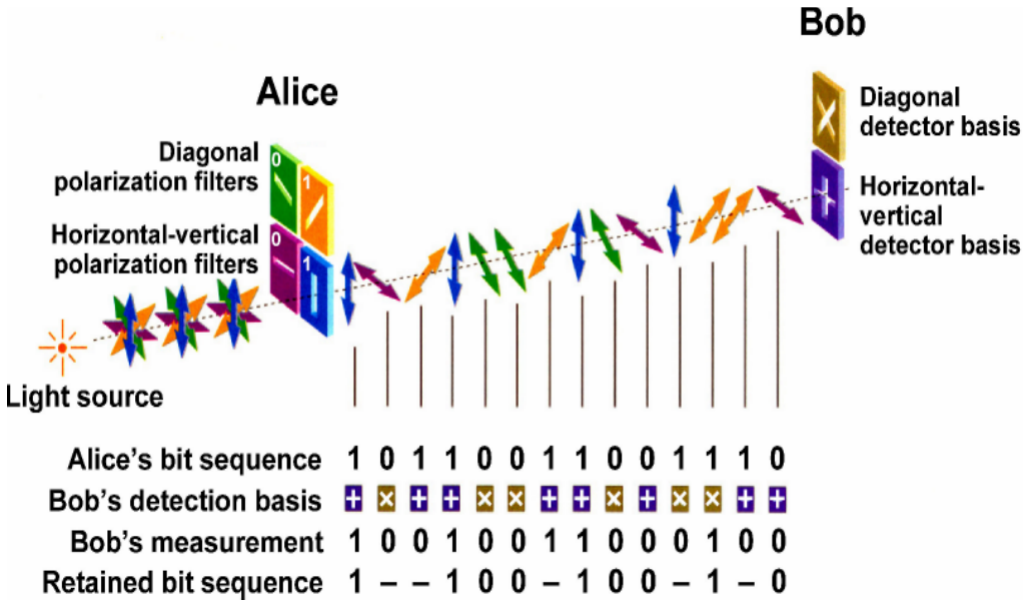

Figure 3.1: Illustration of the BB84 protocol in polarization encoding in the ideal case (perfect system and no eavesdropping). Image courtesy of [50].

After running the protocol, provided that Alice and Bob do not abort, they share a secret key of which Eve has very little knowledge. The key can be used in the one-time-pad cryptosystem for secure communication.

3.1.1(a) The origin of security

The security of BB84 and other QKD protocols can be traced back to several fundamental principles of quantum physics. Since information is encoded in quantum systems unknown to Eve, any attempt by Eve to extract information disturbs the information carriers, which manifests as errors detectable by Alice and Bob. Moreover, the possibility of Eve possessing an identical copy of each quantum signal shared between Alice and Bob is ruled out by the no cloning theorem: it is impossible to perfectly clone an unknown quantum state. Alternatively, in the entanglement based scheme the security can be certified by violation of Bell's inequality because the measurement outcomes do not exist before the measurement, and thereby cannot be created by pre-established agreement (i.e. Eve could not have pre-established Alice and Bob's correlation).

3.1.2 Generic QKD protocol

The BB84 protocol reflects the general structure of any discrete variable QKD protocol. There are two main phases in such a protocol: the signal exchange phase and classical information processing phase. Having agreed upon a common encoding of classical information in quantum systems, Alice and Bob perform signal exchange via an ideal quantum channel which does not disturb the state, and then measurements on these systems to obtain classical results. Then they perform classical information processing, namely sifting, parameter estimation, error correction and privacy amplification, to transform their classical results to a common secret key (or abort if necessary).

QKD protocols are sometimes classified according to how quantum resources are distributed. If Alice prepares the quantum system and sends to Bob for measurement, the protocol is called a quantum key distribution or prepare-and-measure scheme. If Alice and Bob share some entangled quantum state distributed from some source and perform local measurements on their respective part of the joint system, the protocol is called a quantum key distillation or entanglement-based scheme. Any prepare-and-measure scheme has an equivalent entanglement based partner.

For any protocol, we are mainly interested in the number of secure key bits which can be generated. Toward this goal, we must first answer the question: what does it mean for a QKD protocol to be secure?

3.1.2(a) Security requirements

The task of key distribution requires Alice and Bob, at the end of the protocol, to have identical key bits unknown to Eve. To formalize these notions, let E be the quantum system describing the information Eve gathered during the execution of the protocol and K_A, K_B be Alice and Bob's key systems (which are assumed to be on the same key space $\mathcal{K} = \{0,1\}^\ell$). The protocol transforms initial quantum resources to a final classical-quantum state describing Eve's correlation with Alice and Bob's keys

$$\rho_{K_A K_B E} = \sum_{k_A, k_B} P_{K_A K_B}(k_A, k_B) \left|k_A\right\rangle \left\langle k_A\right| \otimes \left|k_B\right\rangle \left\langle k_B\right| \otimes \sigma_E^{k_A, k_B} \tag{3.1}$$

for orthonormal bases $\{|k_A\rangle\}$ and $\{|k_B\rangle\}$. The ideal situation required by the task corresponds to the ideal state

$$\rho_{K_A K_B E}^{\text{ideal}} = \sum_k \frac{1}{2^\ell} \left|k\right\rangle \left\langle k\right| \otimes \left|k\right\rangle \left\langle k\right| \otimes \rho_E \tag{3.2}$$

describing an adversary Eve completely uncorrelated to a uniformly distributed key which are idential for Alice and Bob. In reality, such situation are not possible, so we allow some small failure probability ϵ:

Definition 1. *A QKD protocol is called ϵ-secure if* $\frac{1}{2}\left\|\rho_{K_AK_BE} - \rho^{\text{ideal}}_{K_AK_BE}\right\|_1 \leq \epsilon$.

The reason for this definition is clear from the operational interpretation of the trace distance: the probability of distinguishing the real situation from the ideal situation is at most $1/2 + \epsilon/2$ according to Holevo-Helstrom theorem. Moreover, this security definition is universally composable so any key generated by a QKD protocol can be safely used in any other task such as one-time-pad [51].

In the security analysis of a protocol, it is convenient to break the analysis into an argument about correctness and secrecy.

Definition 2. *(Correctness). A QKD protocol is called ϵ_{EC}-correct if*

$$\Pr[S_A \neq S_B] \leq \epsilon_{\text{EC}}. \tag{3.3}$$

Definition 3. *(Secrecy). A QKD protocol is called ϵ_{sec}-secret if*

$$\frac{1}{2}\left\|\rho_{S_AE} - \rho^{\text{ideal}}_{S_AE}\right\|_1 \leq \epsilon_{\text{sec}}. \tag{3.4}$$

Definition 4. *(Security). A QKD protocol is called ϵ-secure if it is ϵ_{EC}-correct and ϵ_{sec}-secret with $\epsilon_{\text{EC}} + \epsilon_{\text{sec}} < \epsilon$.*

3.1.2(b) Extractable secret key rate

Using the leftover hash lemma an ϵ_{sec}-secret key of length ℓ can be extracted by two-universal hashing if there exists a smoothing parameter $\bar{\epsilon} \geq 0$ such that

$$\ell \leq H^{\bar{\epsilon}}_{\min}(X|E') - 2\log\frac{1}{2(\epsilon_{\text{sec}} - \bar{\epsilon})}. \tag{3.5}$$

where E' represents the information Eve has gathered before privacy amplification. In particular E' contains the information which Eve learns during the error correction phase of the protocol and any additional information E before that. Using a one-way error correcting protocol, we have

$$H^{\bar{\epsilon}}_{\min}(X|E') \geq H^{\bar{\epsilon}}_{\min}(X|E) - \text{leak}_{\text{EC}} - \log\frac{2}{\epsilon_{\text{EC}}} \tag{3.6}$$

where leak_{EC} is the number of bits one party have to send the other and ϵ_{EC} is the probability of failure of the error correction phase (as determined by the

choice of the error correcting code). Hence, we can generate an ϵ-secure key with $\epsilon = \epsilon_{\text{sec}} + \epsilon_{\text{EC}}$ provided

$$\ell \leq H_{\min}^{\bar{\epsilon}}(X|E) - 2\log\frac{1}{2(\epsilon_{\text{sec}} - \bar{\epsilon})} - \text{leak}_{\text{EC}} - \log\frac{2}{\epsilon_{\text{EC}}}. \tag{3.7}$$

Equivalently, a rate $r_N := \ell/N$ can be obtained with

$$r_N = \frac{n}{N}\left(\frac{1}{n}H_{\min}^{\bar{\epsilon}}(X|E) - \frac{2}{n}\log\frac{1}{2(\epsilon_{\text{sec}} - \bar{\epsilon})} - \frac{1}{n}\text{leak}_{\text{EC}} - \frac{1}{n}\log\frac{2}{\epsilon_{\text{EC}}}\right), \tag{3.8}$$

where $N = n + k$ is the total number of signals, n the number of signals devoted for the raw key and k is the number of signals for parameter estimation. Most of the difficulty with a finite key security proof lies in finding a good lower bound for $H_{\min}^{\bar{\epsilon}}(X|E)$ compatible with the observed data, i.e. the error rates. The main reason is that there is not a good characterization of the coherent attacks of Eve or equivalently, the joint state $\rho_{X^N Y^N E}$ after signal exchange phase has no simple structure. This is where techniques such as the de Finetti theorem or the uncertainty relations can be useful.

In the asymptotic limit $N \to \infty$, n/N converges to the sifting probability which can be assumed to converge to 1 because the remaining k signals are still sufficient for a sharp parameter estimation. Moreover, by permuting the classical outcomes and using the de Finetti theorem one can reduce a security statement about coherent attacks to one about collective attacks. It is beyond the scope of this thesis to present the full details of this reduction and we refer to Chapter 6 of [23] for a proof. The asymptotic key rate is given by

$$r_\infty = \min_{\sigma_{AB}} H(X|E) - H(X|Y), \tag{3.9}$$

where the entropies are evaluated on a state σ_{ABE} which is a purification of σ_{AB} and the minimization is over all σ_{AB} compatible with the observed statistics of the different measurements performed by the two parties. From now on, consider only the asymptotic key rate as the benchmark for the performance of a QKD protocol.

3.2 Tomographically complete QKD protocols

3.2.1 Reference frames in QKD

Reference frames play an important role in physics; they are conventions we agree in order to unambiguously define various variables of a physical system.

Unsurprisingly, most QKD protocols implicitly assume a shared reference frame for the quantum communication between the two parties. For instance, in the BB84 protocol, Alice and Bob shares the same understanding of horizontal and vertical directions despite being physically far apart.

Shared reference frame is a resource that should not be taken for granted, however, because the establishment of such requires lots of resources communicated between the corresponding parties [52]. In some scenario, it is even desirable not to try establishing such a shared reference frame because of the natural constraints imposed by the scenario. The first example one may conjure is QKD between an earth station and an orbiting satellite. Apart from a direct communication link between the station and the satellite where circular polarization is unambiguously defined, the linear polarizations may vary in time because the satellite may be rotating with respect to the station. The second example is path-encoded chip-to-chip QKD where the goal is establishing a key between integrated quantum photonic circuits. There it is known that the which-path information is very stable compared to the interferometric stability between Alice and Bob's chips.

One method of performing reference-frame-independent quantum communication is encoding information in decoherence-free subspace or decoherence-free subsystem of a large composite physical system [52]. Suppose ρ, defined with respect to Alice's reference frame is sent to Bob via an ideal quantum channel. Let the unitary operator relating Bob's reference frame to Alice's be $U(\boldsymbol{g})$ parametrized by a set of parameters labeled $\boldsymbol{g}$. Let G be a group of operations relating Bob's frame to Alice's frame; for instance, G can be the group of three dimensional rotations and U can be the unitary representation of this group. If $\boldsymbol{g} \in G$ is unknown, Bob's state with respect to his reference frame reads

$$\tilde{\rho} = \int_G \mathrm{d}\boldsymbol{g} U(\boldsymbol{g}) \rho U^\dagger(\boldsymbol{g}), \tag{3.10}$$

which is completely different from ρ in general. However, there may exit states that are invariant under the action of such G (i.e. $\tilde{\rho} = \rho$), and these span the so-called decoherence-free subspace[1]. The simplest non-trivial example of such decoherence-free subspace is the antisymmetric subspace of two spins 1/2 which is spanned by the singlet (total spin 0) $|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. Without a shared Cartesian frame, Alice can communicate one classical bit to Bob with every two qubits by the following encoding: states in the antisymmetric subspace code for bit 0, while states in the symmetric subspace (orthogonal to the antisymmetric one)

[1] It should be noted that the decoherence free subspace may not exist or may be trivial for a particular system with a particular type of noise.

code for bit 1. Bob then performs a projective measurement onto the symmetric and antisymmetric subspaces defined by his reference frame to determine the bit sent by Alice. To send a qubit, we need a larger composite system, namely three physical spin 1/2.

This kind of encoding has also be applied to QKD. Boileau *et. al* proposed a polarization based protocol where Alice sends photon pairs in state $|\Psi^-\rangle$ and Bob measures the polarization of individual photon [53]. However, as the amount of resources increases so is the sensitivity of the protocol to photon losses. Later, Aolita and Walborn improve the protocol by encoding in the decoherence free subspace of two degrees of freedom of a single photon, namely the polarization and transverse spatial degree of freedom (or transverse spatial profile), therefore solve the above problem [54]. The main drawback with this approach is still the complexity in manipulating multiple systems or multiple degrees of freedom in the same system.

The reference-frame-independent (henceforth abbreviated rfi) QKD protocol tackles the problem of reference frames in a slightly different manner by utilizing the naturally aligned basis (circular polarization or which-path).

3.2.2 Reference frame independent protocols

The main idea in the rfi protocol is that Alice and Bob share a common well-aligned measurement basis $Z = Z_A = Z_B$ which is taken to be the key basis, while the other measurements can be misaligned by an arbitrary but fixed angle β

$$X_B = \cos\beta X_A + \sin\beta Y_A, \quad Y_B = \cos\beta Y_A - \sin\beta X_A. \tag{3.11}$$

As usual, we have the quantum bit error rate in the Z basis,

$$Q = \Pr(a \neq b | Z \otimes Z) = \frac{1 - \langle Z \otimes Z \rangle}{2}, \tag{3.12}$$

where $\langle Z \otimes Z \rangle$ is the average value of the joint measurement Z of both parties. Instead of the error rate in the other basis, the protocol uses the β-independent parameter

$$C = \langle X_A \otimes X_B \rangle^2 + \langle X_A \otimes Y_B \rangle^2 + \langle Y_A \otimes X_B \rangle^2 + \langle Y_A \otimes Y_B \rangle^2 \tag{3.13}$$

to bound Eve's information. C is an entanglement witness: $C \leq 1$ for separable states and $C = 2$ for maximally entangled states.

3.2.2(a) A family of rfi protocols for qudits

We propose a generalization of this protocol to qudits. Let $\{|0\rangle, |1\rangle, ..., |d-1\rangle\}$ be the computational basis vector of the Hilbert space describing a qudit, it is well known that the Pauli operators admit a generalization to higher dimension known as the Weyl operators, which are unitary operators of the form $X^k Z^\ell$ for $k, \ell \in \{0, 1, ..., d-1\}$ and

$$Z = \sum_{j=0}^{d-1} \omega^j |j\rangle \langle j|, \quad X = \sum_{j=0}^{d-1} |j+1\rangle \langle j|, \tag{3.14}$$

where $\omega = e^{2\pi i/d}$ are the roots of unity and $j+1$ denotes the sum modulo d. To accommodate relative unitary rotation around Z (similar to the relative frame angle β in the qubit protocol), let $X_A = UXU^\dagger$ and $X_B = VXV^\dagger$ where $[U, Z] = [V, Z] = 0$. In the generalized protocol, Alice and Bob perform the projective measurements on the eigenstates of $X_A^{k_1} Z^{\ell_1}$ and $X_B^{k_2} Z^{\ell_2}$, and from the statistics estimate Q and

$$C = \sum_{\substack{k_1,k_2=1 \\ \ell_1,\ell_2=0}}^{d-1} |\langle X_A^{k_1} Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2}\rangle|^2 \tag{3.15}$$

to bound Eve's information. (3.15) is a generalization of (3.13) with all the desired properties: it is independent of the local unitaries U and V mentioned above and is an entanglement witness ($C \leq (d-1)^2$ for separable states and $C = d(d-1)$ for maximally entangled states).

The essential ingredients in the proof that equation (3.15) generalizes C are twofold: (i) the relation between average values of operators and the Hilbert-Schmidt inner product, namely $\langle O\rangle_\rho = \langle \rho, O\rangle$ where $\langle \cdot, \cdot \rangle$ is the Hilbert-Schmidt inner product, and (ii) the Weyl operators as an orthonormal basis up to normalization, i.e. $\langle X^{k_1} Z^{\ell_1}, X^{k_2} Z^{\ell_2}\rangle = d\delta_{k_1,k_2}\delta_{\ell_1,\ell_2}$. We recall the computation of inner product using an orthonormal basis

$$\langle \rho, \rho\rangle = \frac{1}{d} \sum_{k,\ell=0}^{d-1} \langle \rho, X^k Z^\ell\rangle \langle X^k Z^\ell, \rho\rangle = \frac{1}{d} \sum_{k,\ell=0}^{d-1} |\langle \rho, X^k Z^\ell\rangle|^2. \tag{3.16}$$

To prove that C is invariant with respect to rotations around Z, first note that C can be rewritten as

$$C = \sum_{\substack{k_1,k_2=0\\ \ell_1,\ell_2=0}}^{d-1} |\langle X_A^{k_1} Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2}\rangle|^2 - \sum_{\substack{k_2,\ell_2=0\\ \ell_1=0}}^{d-1} |\langle Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2}\rangle|^2 - \sum_{\substack{k_1,\ell_1=0\\ \ell_2=0}}^{d-1} |\langle X_A^{k_1} Z^{\ell_1} \otimes Z^{\ell_2}\rangle|^2 + \sum_{\ell_1,\ell_2=0}^{d-1} |\langle Z^{\ell_1} \otimes Z^{\ell_2}\rangle|^2 \tag{3.17}$$

where the first sum simplifies to $d^2 \operatorname{Tr}(\rho_{AB}^2)$. We can switch bases from $X_B^{k_2} Z^{l_2}$ to $X^{k_2} Z^{l_2}$ since they are both bases for the space of linear operators $\mathcal{L}(\mathbb{C}^d)$ on $\mathbb{C}^d$, thus invariant, and similarly for the third sum. The final term is obviously invariant with respect to Z rotations. Therefore, we have proved that C is independent of the local unitaries U and V commuting with Z.

To show that C acts as an entanglement witness, consider the product state $\rho_{AB} = \sigma_A \otimes \sigma_B$ for which C factorizes into

$$C = \sum_{\substack{k_1=1\\ \ell_1=0}}^{d-1} |\langle X_A^{k_1} Z^{\ell_1}\rangle_{\sigma_A}|^2 \sum_{\substack{k_2=1\\ \ell_2=0}}^{d-1} |\langle X_B^{k_2} Z^{\ell_2}\rangle_{\sigma_B}|^2 \tag{3.18}$$

and note that

$$\sum_{\substack{k_1=1\\ \ell_1=0}}^{d-1} |\langle X_A^{k_1} Z^{\ell_1}\rangle_{\sigma_A}|^2 = d \operatorname{Tr}(\sigma_A^2) - 1 - \sum_{\ell_1=1}^{d-1} |\langle Z^{\ell_1}\rangle_{\sigma_A}|^2 \leq d-1, \tag{3.19}$$

from which $C \leq (d-1)^2$ for all product states and moreover for all separable states by convexity. Thus if $C > (d-1)^2$ for a particular state, then the state is entangled, however, the converse, that the state is separable for C less than that value, is not implied. Indeed, entangled states can have $C < (d-1)^2$.

Note that C is a sum over tensor products of operators that do not commute with Z, the raw key basis. The maximum value of C is only achieved for maximally entangled states. The maximum value that can be obtained with a separable state is $(d-1)^2$, therefore there is a gap between separable states and maximally entangled states that scales linearly with d.

3.2.3 RFI protocols are tomographically complete

The way measurement results are used in the rfi protocols is not optimal in the sense that the tomographic information deducible from the measurement statistics

can be used directly, instead of via the parameter C. Estimating C requires the knowledge of $[d(d-1)]^2$ correlators $\langle X_A^{k_1} Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2} \rangle$, and combining these into C discards valuable information that can lead to a tighter bound on Eve's information. In other words, these correlators can directly be used to completely specify the state as

$$\rho_{AB} = \frac{1}{d^2} \sum_{k_1,k_2,\ell_1,\ell_2=0}^{d-1} \langle X_A^{k_1} Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2} \rangle X_A^{k_1} Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2} \tag{3.20}$$

in the measurement bases of Alice $\{X_A^{k_1} Z^{\ell_1}\}$ and Bob $\{X_B^{k_1} Z^{\ell_1}\}$. This follows from the fact that collective attacks are the optimal in general for Eve in this scenario by the quantum de Finetti representation theorem [23, 55], so that we can consider a pure state ρ_{ABE} from which the above ρ_{AB} is the marginal state. This is also the reason why the rfi protocols are actually tomographic in disguise: we are trying to do tomography on Eve's optimal state in her collective attack. Therefore by using tomography, one can have a rfi protocol without the need for C [56]. In fact, the correlators can be used to realign the reference frames during the execution of the protocol, if necessary.

Let us explain in detail how tomography can be done in practice. The most direct approach is to make d^2 measurements $X^k Z^\ell$ on each subsystem of each party and combine the measurement outcomes to find each correlator directly. This is very inefficient because it requires d^2 different estimates to be made with good precision, which requires many copies of the state. Also, it is unnecessary because many of the Weyl operators have the same set of eigenvectors, for instance Z^ℓ for $\ell = 0, ..., d-1$ for instance; hence the measurement statistics of one can be used to calculate the average values of all the others. In general, the minimum number of measurements needed to completely specify the state is still unknown. However, if d is prime, one can reconstruct the state by making only $d+1$ measurements corresponding to $d+1$ mutually unbiased bases on each subsystem, the mutually unbiased bases generated by the set of operators $\mathcal{B} = \{Z, XZ^\ell : \ell = 0, ..., d-1\}$ for example. After the measurements, Alice and Bob can estimate their marginal probability distribution locally, and if they share the measurement outcomes, the joint probability distribution $p(a, b|A \otimes B)$ where $A, B \in \mathcal{B}$. It is well known that the eigenbasis of any $X^k Z^\ell$ is among the eigenbases of observables in $\mathcal{B}$; therefore from $p(a, b|A \otimes B)$ we can compute all the average values using

$$\langle X_A^{k_1} Z^{\ell_1} \otimes X_B^{k_2} Z^{\ell_2} \rangle = \sum_{a,b} \lambda_a \lambda_b p(a, b|A \otimes B), \tag{3.21}$$

where λ_a is the eigenvalue associated to the eigenvector representing outcome a of

$X_A^{k_1} Z^{\ell_1}$ and A is the operator in $\mathcal{B}$ with the same eigenbasis as $X_A^{k_1}$, ditto for Bob. Hence, a full state reconstruction is possible by (3.20).

As a side remark, we note that state reconstruction can also be done if a SIC-POVM (symmetric and informationally complete positive operator valued measure) exists for the given dimension. For instance, such a measurement exists for a qubit: the POVM elements are projectors pointing towards the corners of a tetrahedron in the Bloch sphere. However, the implementation of such measurements may be complicated.

3.2.4 Conclusions

Our effort to generalize the rfi QKD for qubits, which arises naturally in several realistic applications, to higher dimension have been hit with the realization that the protocols are actually tomographic in nature. In other words, our family of d dimensional rfi protocols can be seen as a generalization of the six-state protocol. Thus using directly the tomographic information gives a better constraint on the state shared by the users and the adversary, which ultimately gives better key rate. The reference frame independent property of a QKD protocol is not a consequent of the invariance of the parameters (such as C) used in the protocol.

3.3 Distributed-phase-reference QKD

3.3.1 Motivations

When experimentalists try to implement discrete variable protocols (BB84, six-state, or the one presented in the previous section) using quantum optics, they invent a whole new class of QKD protocols called the distributed-phase-reference (DPR) protocols. The major distinction between discrete variable and DPR protocols lie in their way of encoding information. In discrete variable protocols each symbol is encoded in a quantum state distinct from the quantum state encoded any other symbol, while in DPR protocols each symbol is encoded in consecutive pairs of quantum states (laser pulses). The two most well known DPR protocols are the differential phase shift (DPS) and the coherent one way (COW) protocols.

In the DPS protocol [57], Alice sends a sequence of coherent states with the same intensity, and modulates the phase between successive pulses between 0 to code for bit "0" and π to code for bit "1". On Bob's side, he can unambiguously discriminate the encoded bit by interfering successive pulses with an unbalanced interferometer. More specifically, Bob can calibrate his interferometer such that the path length difference makes up for the delay between the pulses, and whenever

their relative phase is 0 then detector D0 will click (likewise for relative phase π and detector D1).

In the COW protocol [58], Alice sends a sequence of empty and non-empty coherence states with the same intensity, and encodes each bit in successive pairs: bit 0 is encoded in the sequence empty, non-empty while bit 1 is encoded in the reverse sequence of non-empty followed by empty pulse. Bob can unambiguously decode each bit by measuring the time of arrival (or time of detection) of each pulse. The security can be guaranteed by sending decoy sequence consisting of two non-empty pulses and check for their relative phases like DPS.

It is clear that in both protocols, there is no clear distinction which pair of pulses encodes for which bit, and therefore the *entire* chain of pulses must be treated as a single, albeit very huge, signal. This hinders the development of a complete security proof of DPR-QKD in a realistic setting. However, security has been proven against restricted types of attacks [36, 37], or assuming single photon sources [38].

In the remaining of this section, we will prove the security of a variant of COW. This variant is the subject of an experiment in the group of Nicolas Gisin. The basic setup is shown in Figure 3.2. Alice uses a laser, followed by an intensity

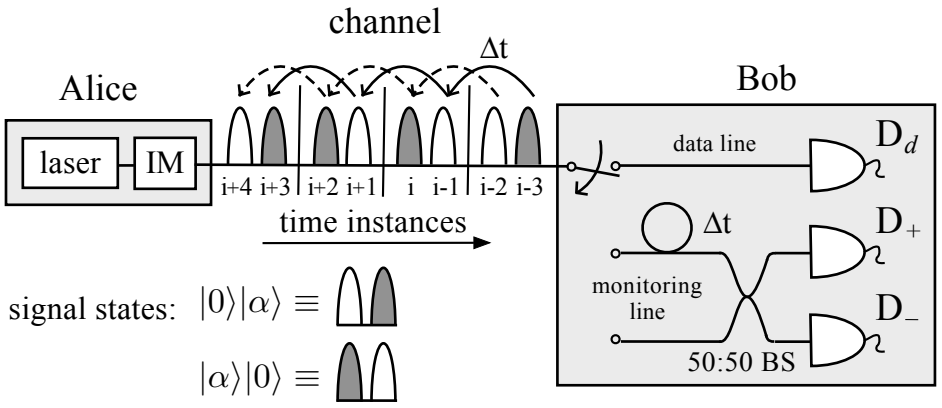

Figure 3.2: Schematic description of a modified version of the COW protocol with an active measurement choice. Bob reads the raw key in detector D_d. Moreover, he uses an optical switch to send some pairs of consecutive pulses to a monitoring line that examines the coherence between even and odd pulses.

modulator (IM), to prepare a sequence of coherent states $|0\rangle\,|\alpha\rangle$ and $|\alpha\rangle\,|0\rangle$. On the receiving side, Bob employs an active optical switch to distribute each pair of incoming pulses into the data or the monitoring line. The data line measures the arrival time of the pulses in detector D_d and creates the raw key. Whenever Bob sees a click in this detector in say time instance i, he decides at random whether to publicly *announce* a detection event in time instances i and $i+2$ or i and $i-2$. The

first case is associated with a bit value "0", while the second corresponds to a bit value "1". If the state sent by Alice in these time instances is $|0\rangle|\alpha\rangle$ (respectively $|\alpha\rangle|0\rangle$) then she assigns to it a bit value "0" (respectively "1") and tells Bob to keep his result. Otherwise, the result is discarded[2]. This public announcement by Alice and Bob is labeled as v later in the text. The monitoring line checks for eavesdropping by measuring the coherence between subsequent even and odd pulses. This is done by interfering adjacent pairs of pulses in a $50:50$ beamsplitter and measuring the outputs in detectors D_+ and D_-.

3.3.2 A framework to security of DPR

3.3.2(a) The de Finetti approach to security

The challenge in DPR-QKD is to prove security against coherent attacks. Usually, such attacks in the discrete variable protocols are known to be of no advantage to Eve in comparison to collective attacks, by virtue of the de Finetti theorem [23, 55]. This theorem applies, for instance, when the underlying quantum state shared by Alice and Bob is permutationally invariant, which is typically ensured by performing simultaneous random permutations on the classical measurement results. DPR-QKD defines however a fixed ordering of the signals by its coherence measurement and, therefore, it is not possible to permute the classical outcomes without destroying vital information. However, such a predicament can be circumvented by grouping the entire signal stream into blocks. More specifically, consider that Alice and Bob group their signals into subsequent blocks of size m with m being optimized for the expected behavior. When permuting these blocks, one preserves the coherence information within them, while the information between the blocks is destroyed. Still, this is enough to *apply the de Finetti argument on the level of blocks*. As a result, the state shared by all parties after distributing a large number mN of signals satisfies $\rho_{\mathrm{ABE}}^{mN} \approx \rho_{\mathrm{ABE}}^{m\,\otimes N}$, and security against collective attacks on these signal blocks implies security against coherent attacks in the original setting.

3.3.2(b) The asymptotic key rate formula

The state in collective attacks shared by all parties after transmitting an m block signal ρ_{ABE}^{m} is not arbitrary and can be related to Alice's preparation procedure, where she sends potentially mixed states ρ_i^m with a priori probability $p(i)$. In the equivalent entanglement based version, Alice first creates a source state $|\Psi^m\rangle_{\mathrm{A_bA_sB}} =$

[2]Note that the two-way communication here is needed to establish the raw key and is not part of the classical post-processing which is chosen to be one-way communication (say from Alice to Bob).

$\sum_i \sqrt{p(i)}\,|i\rangle_{\mathrm{A_b}}\,|\rho_i^m\rangle_{\mathrm{A_sB}}$, where $|\rho_i^m\rangle_{\mathrm{A_sB}}$ denotes purifications of the signal states ρ_i^m to an inaccessible (to Eve) shield system $\mathrm{A_s}$. Afterwards, she measures her bit system $\mathrm{A_b}$ in the standard basis, thereby producing the correct signal states at site B which are sent to Bob. Eve transforms the overall source state to the aforementioned tripartite state ρ^m_{ABE} with $\mathrm{A} = \mathrm{A_bA_s}$. In other words, Eve is allowed to only interact with the system travelling to Bob, hence the collective attack state ρ^m_{ABE} obeys certain constraints of the corresponding DPR protocol (such as fixed marginal on Alice).

To derive the asymptotic key rate, let us first consider the effect of public announcements by Alice and Bob based on their classical measurement results. This announcement, labeled as v, allows both parties to distinguish between conclusive events that contribute to the sifted key and inconclusive ones that are discarded. On the level of quantum states this is described by suitable maps $\Lambda_v^{\mathrm{A}} \otimes \Lambda_v^{\mathrm{B}}$. Given an announcement v that happens with probability $p(v)$, the three parties share the state $\sigma^m_{\bar{\mathrm{A}}\bar{\mathrm{B}}\mathrm{E},v}$ determined by $\Lambda_v^{\mathrm{A}} \otimes \Lambda_v^{\mathrm{B}}(\rho^m_{\mathrm{ABE}}) = p(v)\sigma^m_{\bar{\mathrm{A}}\bar{\mathrm{B}}\mathrm{E},v}$.

For each announcement v one can use the one-way classical post-processing key rate formula [19] (or one can derive it from our asymptotic key rate formula presented before). If system $\bar{A}$ denotes a qubit and Alice's raw key is obtained by projecting this system onto $|0\rangle_{\bar{\mathrm{A}}}, |1\rangle_{\bar{\mathrm{A}}}$, then a lower bound on the secret key rate is given by $1 - h_2(e_v) - h_2(\delta_v)$ with $h_2(p) = -p\log(1-p) - (1-p)\log(1-p)$ the binary entropy associated with the distribution $\{p, 1-p\}$. Here e_v is the symmetrized bit error between the key measurements of Alice and Bob, and δ_v denotes the corresponding error, typically called phase error, when Alice performs a measurement in a mutually unbiased basis and Bob in his other setting. This last parameter is used to upper bound Eve's knowledge on the sifted key generated by Alice. Note that δ_v does not need to be measured directly, it only needs to be estimated.

To consider that the output system $\bar{A}$ is a qubit implies that Alice can, at best, distill one secret bit per block. Nevertheless this restriction should not have a significant impact on the key rate in a long distance regime, since Bob observes, if any, most often only one single conclusive event per m arriving signals due to the high losses in the channel (given that m is not too big).

Instead of estimating separate phase errors δ_v, it is often easier to combine all conclusive announcements $v \in \mathcal{V}_{\mathrm{c}}$ into an averaged version. Let $G = \sum_{v\in\mathcal{V}_{\mathrm{c}}} p(v) \leq 1$ denote the total sifted key gain. Then, we have that the secret key rate per block

is bounded by

$$R_m \geq \inf_{\rho^m_{\mathrm{ABE}}} \sum_{v\in\mathcal{V}_c} p(v)\left[1-h_2(e_v)-h_2(\delta_v)\right] \tag{3.22}$$

$$\geq \inf_{\rho^m_{\mathrm{ABE}}} G\left[1-h_2(\bar{e}_c)-h_2(\bar{\delta}/G)\right]$$

$$\geq G\left[1-h_2(\bar{e}_c)-h_2(\bar{\delta}^{\max}/G)\right]. \tag{3.23}$$

Here one uses concavity of binary entropy to lower bound R_m by the averaged (conditional) error rates $\bar{e}_c = \sum_{v\in\mathcal{V}_c} p(v)e_v/G$ and $\bar{\delta} = \sum_{v\in\mathcal{V}_c} p(v)\delta_v$. The last step takes into account that $\bar{e}_c$ and G are observed quantities and that the optimization is attained at the largest phase error $\bar{\delta}^{\max}$ compatible with the obtained data since h_2 increases in $[0, \frac{1}{2}]$.

3.3.2(c) Phase error estimation SDP program

The main difficulty to compute (3.23) is to upper bound the average phase error $\bar{\delta}$. This parameter can be expressed as an expectation value on the original bipartite state $\rho^m_{\mathrm{AB}} = \mathrm{tr}_{\mathrm{E}}(\rho^m_{\mathrm{ABE}})$ using adjoint maps

$$\bar{\delta} = \sum_{v\in\mathcal{V}_c} p(v)\,\mathrm{tr}(\sigma^m_{\mathrm{A\bar{B}},v}F_{\delta_v}) = \sum_{v\in\mathcal{V}_c} \mathrm{tr}[\Lambda^{\mathrm{A}}_v\otimes\Lambda^{\mathrm{B}}_v(\rho^m_{\mathrm{AB}})F_{\delta_v}]$$

$$= \mathrm{tr}[\rho^m_{\mathrm{AB}}\sum_{v\in\mathcal{V}_c}\Lambda^{\mathrm{A}\dagger}_v\otimes\Lambda^{\mathrm{B}\dagger}_v(F_{\delta_v})] = \mathrm{tr}(\rho^m_{\mathrm{AB}}F_{\bar{\delta}}). \tag{3.24}$$

Here F_{δ_v} denotes the corresponding phase error operators on the state $\sigma^m_{\mathrm{A\bar{B}},v}$. Partial knowledge of Alice and Bob about the state ρ^m_{AB} can be parsed as known expectation values $k_i = \mathrm{tr}(\rho^m_{\mathrm{AB}}K_i)$ for certain operators K_i. More precisely, since on the receiving side Bob performs a measurement modeled by B_k, as a result, both Alice and Bob observe the expectation values of $|i\rangle_{\mathrm{A_b}}\langle i|\otimes\mathbb{1}_{\mathrm{A_s}}\otimes B_k$. Moreover, since Eve is restricted to interact only with Bob's system, the reduced density matrix $\rho^m_{\mathrm{A}} = \mathrm{tr}_{\mathrm{BE}}(\rho^m_{\mathrm{ABE}})$ is fixed and given by the source state. This information can be added by including expectation values of $T_k\otimes\mathbb{1}_{\mathrm{B}}$, where T_k denotes a tomographic complete operator set on A. Both sets of observables constitute the previously denoted K_i. This means that the search for the maximum phase error $\bar{\delta}^{\max}$ can be cast into a semidefinite program,

$$\begin{aligned} \max \quad & \mathrm{tr}(\rho^m_{\mathrm{AB}}F_{\bar{\delta}}) \\ \text{s.t.} \quad & \rho^m_{\mathrm{AB}}\succeq 0,\ \mathrm{tr}(\rho^m_{\mathrm{AB}}K_i) = k_i\ \forall i. \end{aligned} \tag{3.25}$$

Such special convex optimization problems can be solved efficiently using standard tools to obtain the exact optimum, even for large dimensions.

3.3.2(d) Finite dimensionality and the de Finetti theorem

The signal states and performed measurements in practical DPR-QKD are described by operators on an infinite dimensional Fock space of several modes. In order to apply the de Finetti argument [23, 55], and to numerically obtain an upper bound using (3.25), it is necessary to formulate this problem in a manageable, finite dimensional form. Clearly, system A_b is finite dimensional. For Bob's measurements one can employ the squash model argument [59, 60]. Here the real measurement is notionally decomposed into a two step procedure by first applying a map that transforms any incoming signal to a finite dimensional output state on which a specified target measurement B_k is performed afterwards. Since this map can be even given to Eve, its output state only lowers the key generation capabilities of Alice and Bob, and one readily works in finite dimensions. For the shield system $\mathrm{A_s}$ one uses only partial information of the reduced state. In the case of phase randomized signal blocks, an example that we consider later, a purification stores the total number of photons of the block in the shield system $|n\rangle_{\mathrm{A_s}}$. Using tomography on the subspace spanned by all $n = 1, \ldots, n_\mathrm{cut}$, together with an ancilla $|N\rangle_{\mathrm{A_s}}$ for all other cases, the shield system can effectively be described in finite dimensions.

3.3.3 Security analysis of a variant of COW

For the described COW protocol we consider blocks carrying m bits of information. Since a single bit comprises two modes, one has $2m$ different temporal modes described by their creation and annihilation operators $a_s^\dagger$ and a_s, respectively, with $s = 1, \ldots, 2m$. We assume that the l-th bit relates to the modes with $s = 2l - 1, 2l$. In the security analysis we assume that Alice and Bob discard coherence information between consecutive blocks and that the sifted key is created only from signals within the same block. To guarantee this, one could discard detection events where Bob declares time instances that belong to different blocks.

3.3.3(a) Real and assumed measurement description

At first let us concentrate on the real measurement model M_k^real and the way how we describe it in the security part, denoted as B_k previously. For the real measurement setup we assume inefficient photon number resolving detectors that suffer from state-independent dark counts. The inefficiency of M_k^real is modeled by a global

beamsplitter (BS) of transmittance η_{det} located in front of a perfectly efficient scheme, labeled as M_k, that still suffers from dark counts. This is schematically drawn in the first line of Figure 3.3. In a second step, one models the efficient scheme M_k as a map Λ_s, sometimes called squashing or filter operation [59], in front of the assumed description B_k. Let us emphasize that the security simulation is valid for any true measurement scheme that can be modeled as a physical map Λ followed by the measurement B_k as shown in the third line of the figure.

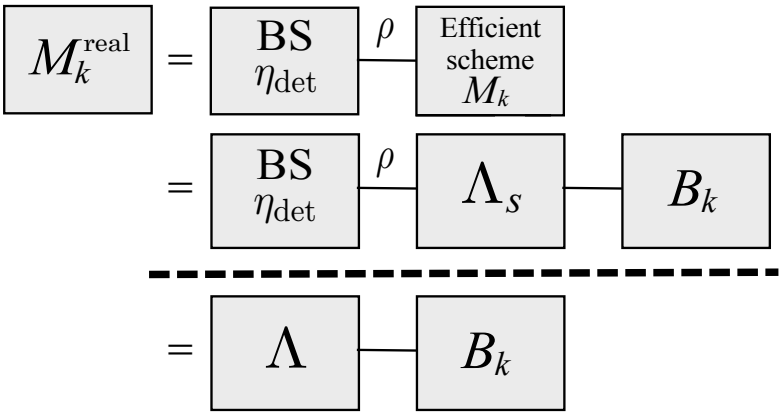

Figure 3.3: Decomposition of Bob's measuring device.

There are three different types of outcomes for the so far abstract outcome label "k". (i) For a data line measurement we use d, with $d = 1, \ldots, 2m$, to denote a *single photon detection* in temporal mode d only. The corresponding measurement operator M_d is given by

$$M_d = \epsilon(1-\epsilon)^{2m-1} |\text{vac}\rangle \langle\text{vac}| + (1-\epsilon)^{2m} |d\rangle \langle d| , \tag{3.26}$$

with ϵ representing the dark count probability of Bob's detectors and $|d\rangle = a_d^\dagger |\text{vac}\rangle$. (ii) In addition to a data line measurement Bob can also perform coherence measurements on subsequent bits employing the monitoring line. For instance, whenever he tests the coherence between bits l and $l+1$ he effectively mixes the modes $2l-1, 2l+1$ and, at the same time, $2l, 2l+2$. For each pair of modes there are two single photon events, denoted as $\pm$, that can be distinguished, depending on whether the single excitation is registered in the bright (D_+) or in the dark (D_-) detector. As an outcome label for the coherence measurements we use $k = (c, \pm)$, where $c = 1, \ldots, 2m-2$ denotes the first of the two interfering modes. In this case the measurement operators are given by

$$M_{c,\pm} = \epsilon(1-\epsilon)^{2m-1} |\text{vac}\rangle \langle\text{vac}| + (1-\epsilon)^{2m} \left|\chi_c^\pm\right\rangle \left\langle\chi_c^\pm\right| , \tag{3.27}$$

with $|\chi_c^{\pm}\rangle = (|c\rangle \pm |c+2\rangle)/\sqrt{2}$. Let us emphasize that in these coherence measurements it is still necessary to check that all other modes are empty. (iii) Finally, note that each measurement setting has also other possible outcomes, *e.g.*"no click" or more than a single photon detection event. All these cases are grouped (via classical post-processing) into a single inconclusive outcome described by M_{inc}.

As the modeled measurement operators B_k we use

$$B_d = |d\rangle\langle d|\,, \tag{3.28}$$

$$B_{c,\pm} = \left|\chi_c^{\pm}\right\rangle\left\langle\chi_c^{\pm}\right|, \tag{3.29}$$

$$B_{\text{inc}} = |a\rangle\langle a|\,, \tag{3.30}$$

where $|a\rangle$ is the auxiliary state that describes the inconclusive outcome. These measurement operators B_k act on a $2m+1$ dimensional Hilbert space.

Both measurement sets can be made equivalent by an appropriate map Λ_s such that $\text{tr}(\rho M_k) = \text{tr}[\Lambda_s(\rho)B_k]$ holds for all possible states ρ and measurement outcomes "k" as schematically shown in Figure 3.3. This map Λ_s is given as follows. First one measures the total number of photons n within an arriving block. Whenever one finds $n \geq 2$ one outputs the auxiliary state $|a\rangle$. If $n = 1$ then with probability $(1-\epsilon)^{2m}$ the single photon state stays untouched, otherwise the auxiliary state is thrown again. Finally, for $n = 0$ the map creates the completely mixed single photon state $\sum_k |k\rangle\langle k|/2m$ with probability $2m\epsilon(1-\epsilon)^{2m-1}$ and $|a\rangle$ otherwise. This map is physical because we explicitly describe it in terms of measurements and conditional signal state preparations.

3.3.3(b) Source state and reduced density matrix

The following discussion provides the source states for both cases of pure or phase randomized COW block signals. These states determine the reduced density matrix ρ_{A}^m which belongs to the available information.

Consider first the case of pure signal states. In this COW version Alice sends to Bob either the sequence $|\alpha, 0\rangle$ or $|0, \alpha\rangle$, with $\alpha \in \mathbb{R}$, depending on whether her raw key bit value is "0" or "1". Let us start with the scenario where Alice sends to Bob only one bit value, occurring with equal a priori probability. This corresponds to a block size $m = 1$. Then the source state is given by

$$\left|\Psi^{m=1}\right\rangle_{\text{AB}} = \frac{1}{\sqrt{2}}\Big(|0\rangle_{\text{A}}|\alpha,0\rangle_{\text{B}} + |1\rangle_{\text{A}}|0,\alpha\rangle_{\text{B}}\Big), \tag{3.31}$$

and its reduced density matrix becomes

$$\rho_{\mathrm{A}}^{m=1} = \frac{1}{2}\begin{bmatrix} 1 & e^{-\alpha^2} \\ e^{-\alpha^2} & 1 \end{bmatrix}. \tag{3.32}$$

Suppose now that Alice sends to Bob m bits according to this scheme. If $i = (i_1, i_2, \ldots, i_m)$ denotes the m-bit string being sent and $|\phi_i\rangle_{\mathrm{B}}$ refers to the corresponding signal state, then one obtains

$$\begin{aligned} |\Psi^m\rangle_{\mathrm{AB}} &= 2^{-\frac{m}{2}} \sum_{i\in\{0,1\}^m} |i\rangle_{\mathrm{A}} |\phi_i\rangle_{\mathrm{B}} \\ &= |\Psi^m\rangle_{\mathrm{A_1 \ldots A_m B}} = \left|\Psi^{m=1}\right\rangle_{\mathrm{AB}}^{\otimes m}. \end{aligned} \tag{3.33}$$

In particular, from the last expression one finds that the reduced density matrix ρ_{A}^m is given by

$$\rho_{\mathrm{A}}^m = (\rho_{\mathrm{A}}^{m=1})^{\otimes m}. \tag{3.34}$$

Next, let us turn to the case of phase randomized blocks. Since randomizing the phase of a block is equivalent to measuring the total number of photons contained in it, the true signals states are of the form

$$\rho_i^m = \sum_{n=0}^{\infty} \Pi_n |\phi_i\rangle_{\mathrm{B}} \langle\phi_i| \Pi_n = \sum_{n=0}^{\infty} p_\lambda(n) \left|\psi_n^i\right\rangle_{\mathrm{B}} \left\langle\psi_n^i\right|. \tag{3.35}$$

Here Π_n stands for the projector onto the n-photon subspace of the $2m$ different modes. The outcome of such a photon number measurement follows a Poisson distribution $p_\lambda(n) = e^{-\lambda}\lambda^n/n!$ with mean $\lambda = m\alpha^2$. The projected n-photon signal states $|\psi_n^i\rangle_{\mathrm{B}}$ can be expressed as

$$\left|\psi_n^i\right\rangle_{\mathrm{B}} = m^{-\frac{n}{2}}\sqrt{n!} \sum_{n_1,\ldots n_m} \prod_{l=1}^{m} \frac{(a_{2l+i_l-1}^\dagger)^{n_l}}{n_l!} |\mathrm{vac}\rangle_{\mathrm{B}}, \tag{3.36}$$

where the summation runs over all natural numbers $n_1, \ldots, n_m$ that satisfy $\sum_{l=1}^m n_l = n$. These states fulfill the relation

$$\langle\psi_n^i|\psi_{\bar{n}}^j\rangle = \delta_{n\bar{n}} \left(\frac{m - \Delta_{ij}}{m}\right)^n, \tag{3.37}$$

with Δ_{ij} being the Hamming distance between the bit strings i and j, i.e. the number of places they differ.

Using the framework of mixed signal states as explained in the last section one must now choose an overall purification of all signal states $|\psi_n^i\rangle_{\mathrm{B}}$ to a shield system

$|\rho_i^m\rangle_{\mathrm{A_sB}}$. However let us point out that though a single purification $|\rho_i^m\rangle_{\mathrm{A_sB}}$ is unique up to local unitary, here one requires that all signals ρ_i^m are purified to the same shield A_s, which is not unique anymore. While certain collective purifications are clearly better than others, any choice is valid. For our simulation we select

$$|\rho_i^m\rangle_{\mathrm{A_sB}} = \sum_{n=0}^{\infty} \sqrt{p_\lambda(n)}\, |n\rangle_{\mathrm{A_s}} \left|\psi_n^i\right\rangle_{\mathrm{B}}, \tag{3.38}$$

which can be seen as a coherent storage of the total photon number n in the shield system $\mathrm{A_s}$. Let us remark that this choice satisfies $\langle\rho_i^m|\rho_j^m\rangle = F(\rho_i^m, \rho_j^m)$, with F being the fidelity of mixed states, which is also the maximal possible overlap between two signal states [45]. We find, therefore, that the source state in this scenario is given by

$$|\Psi^m\rangle_{\mathrm{A_bA_sB}} = 2^{-\frac{m}{2}} \sum_{i\in\{0,1\}^m} |i\rangle_{\mathrm{A_b}} |\rho_i^m\rangle_{\mathrm{A_sB}}, \tag{3.39}$$

with $\mathrm{A_b} = \mathrm{A_1} \ldots \mathrm{A_m}$. This means that the reduced density matrix ρ_A^m, with $\mathrm{A} = \mathrm{A_bA_s}$, can be expressed as

$$\rho_\mathrm{A}^m = \sum_{n=0}^{\infty} p_\lambda(n) \rho_{\mathrm{A_b}}^n \otimes |n\rangle_{\mathrm{A_s}} \langle n|, \tag{3.40}$$

with $\rho_{\mathrm{A_b}}^n$ given by

$$\rho_{\mathrm{A_b}}^n = 2^{-m} \sum_{i,j} \left(\frac{m - \Delta_{ij}}{m}\right)^n |i\rangle_{\mathrm{A_b}} \langle j|. \tag{3.41}$$

In our simulation we only use partial information of the reduced density matrix ρ_A^m. In particular, we transform $\mathrm{A_s}$ to $\bar{\mathrm{A}}_\mathrm{s}$ by making a shield measurement that distinguishes the different photon number cases mentioned in the main text such that one obtains

$$\begin{aligned}\rho_{\mathrm{A_b}\bar{\mathrm{A}}_\mathrm{s}}^m = &\sum_{n=1}^{n_\mathrm{cut}} p_\lambda(n) \rho_{\mathrm{A_b}}^n \otimes |n\rangle_{\bar{\mathrm{A}}_\mathrm{s}} \langle n| \\ &+ \sum_{n\notin\{1,\ldots,n_\mathrm{cut}\}} p_\lambda(n) \rho_{\mathrm{A_b}}^n \otimes |N\rangle_{\bar{\mathrm{A}}_\mathrm{s}} \langle N|,\end{aligned} \tag{3.42}$$

where $|N\rangle_{\bar{\mathrm{A}}_\mathrm{s}}$ denotes an auxiliary system for all higher photon numbers. Let us point out that considering the reduced state given by (3.42) can be understood as "tagging" the $n = 1, \ldots, n_\mathrm{cut}$ signal states [61].

3.3.3(c) Announcement maps and phase operator

The specific announcements v of the COW protocol can be phrased in terms of appropriate maps Λ_v on the quantum state. Together with a chosen "phase setting" measurement this provides a concrete expression for the averaged phase error operator $F_{\bar{\delta}}$ used in (3.24).

As explained in the protocol description, Bob announces two consecutive even or odd time slots where he registered his single photon event. Suppose, for instance, that he announces $v = (2l-1, 2l+1)$. These are the first arrival times of the modes associated with bits i_l and i_{l+1} sent by Alice. In such cases, Alice and Bob agree to call the outcome in the first time instance "0" while the later event is "1". This announcement can be modeled as a filter operation $\Lambda_v^{\mathrm{B}}(\rho) = F_v^{\mathrm{B}} \rho F_v^{\mathrm{B}\dagger}$ given by

$$F_v^{\mathrm{B}} = \frac{1}{\sqrt{2}} \left(|0\rangle_{\bar{\mathrm{B}}\mathrm{B}} \langle 2l-1| + |1\rangle_{\bar{\mathrm{B}}\mathrm{B}} \langle 2l+1| \right). \tag{3.43}$$

If Bob measures system $\bar{\mathrm{B}}$ in the standard basis $|0\rangle_{\bar{\mathrm{B}}}, |1\rangle_{\bar{\mathrm{B}}}$ he obtains the real outcome he has observed. The pre-factor $1/\sqrt{2}$ which appears in (3.43) takes into account that whenever Bob sees a single photon click in either $2l-1$ or $2l+1$ he announces this particular v with just 50% probability, i.e. $F_v^{\mathrm{B}\dagger} F_v^{\mathrm{B}} = (B_{2l-1} + B_{2l+1})/2$.

Suppose Bob has actually declared $v = (2l-1, 2l+1)$. Then, Alice has to look on her bit string to determine whether she can conclusively infer Bob's bit value. For that, only her bits i_l and i_{l+1} matter. As shown in Figure 3.4, if these two bits are equal it means that she had sent to Bob either two full or two empty pulses. In this scenario, she cannot infer Bob's bit value and they discard this result. However, if these bits differ then she knows Bob's sifted bit value precisely (in the error free case) and she tells Bob to keep it. Such a conclusive announcement by Alice

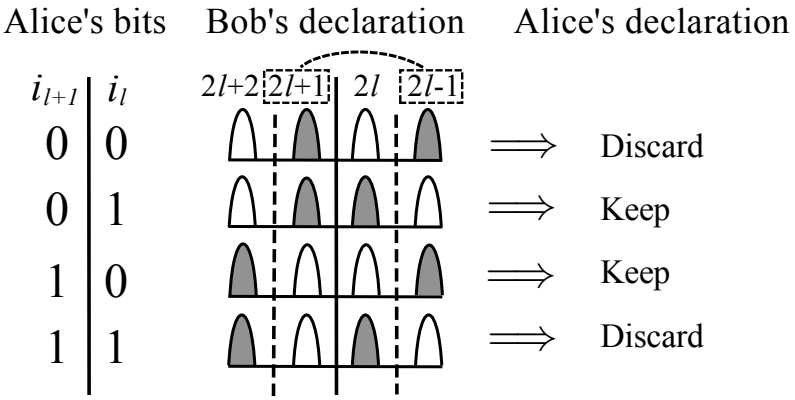

Figure 3.4: Announcement choices for Alice given that Bob has declared a detection event in time slots $2l-1$ and $2l+1$.

can similarly be modeled as a filter operation Λ_v^{A} acting on her qubits l and $l+1$, *i.e.*$\Lambda_v^{\mathrm{A}}(\rho^m_{\mathrm{A}_1\ldots\mathrm{A}_m}) = F_v^{\mathrm{A}}\rho^m_{\mathrm{A}_l\mathrm{A}_{l+1}}F_v^{\mathrm{A}\dagger}$ with

$$F_v^{\mathrm{A}} = |0\rangle_{\bar{\mathrm{A}}\ \mathrm{A}_l\mathrm{A}_{l+1}}\langle 01| + |1\rangle_{\bar{\mathrm{A}}\ \mathrm{A}_l\mathrm{A}_{l+1}}\langle 10|\,. \tag{3.44}$$

Again a measurement in the standard basis $|0\rangle_{\bar{\mathrm{A}}}, |1\rangle_{\bar{\mathrm{A}}}$ provides Alice with her real outcomes.

In order to determine the phase error δ_v we assume that both parties perform measurements in the X-basis, i.e. they project the output signals from their filter operations onto the states $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Then, the symmetrized phase error $\delta_v = p(+,-) + p(-,+)$ can be expressed as

$$\begin{aligned} p(v)\delta_v &= p(v)\,\mathrm{tr}\left[\frac{1}{2}(\mathbb{1}\otimes\mathbb{1} - \sigma_x\otimes\sigma_x)\sigma^m_{\bar{\mathrm{A}}\bar{\mathrm{B}},v}\right] && (3.45)\\ &= \frac{1}{2}p(v) - \frac{1}{2}\,\mathrm{tr}(\sigma_x\otimes\sigma_x p(v)\sigma^m_{\bar{\mathrm{A}}\bar{\mathrm{B}},v}) && (3.46)\\ &= \frac{1}{2}p(v) - \mathrm{tr}(X'_{\mathrm{A}}\otimes X'_{\mathrm{B}}\rho^m_{\mathrm{AB}}), && (3.47)\end{aligned}$$

with σ_x denoting the Pauli matrix $\sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0|$. In (3.47) we have defined the operators

$$\begin{aligned} X'_{\mathrm{A}} &= \mathbb{1}_{\mathrm{A}_1\ldots\mathrm{A}_{l-1}}\otimes X_{\mathrm{A}}\otimes\mathbb{1}_{\mathrm{A}_{l+2}\ldots\mathrm{A}_m}, && (3.48)\\ X'_{\mathrm{B}} &= \frac{1}{2}F_v^{\mathrm{B}\dagger}\sigma_x F_v^{\mathrm{B}} = \frac{1}{4}\left(|2l-1\rangle\langle 2l+1| + |2l+1\rangle\langle 2l-1|\right), && (3.49)\end{aligned}$$

with $X_{\mathrm{A}} = F_v^{\mathrm{A}\dagger}\sigma_x F_v^{\mathrm{A}} = |01\rangle\langle 10| + |10\rangle\langle 01|$.

Similar arguments apply to the cases where Bob announces subsequent even outcome pairs or the special instances at the borders of the blocks. We find that the averaged phase error $\bar{\delta} = \sum_{v\in\mathcal{V}_{\mathrm{c}}} p(v)\delta_v$ can be written as

$$\bar{\delta} = \frac{1}{2}\sum_{v\in\mathcal{V}_{\mathrm{c}}} p(v) - \mathrm{tr}(X_{\bar{\delta}}\rho^m_{\mathrm{AB}}), \tag{3.50}$$

with an operator $X_{\bar{\delta}} = \sum_{l=1}^m X_{\mathrm{A};l}\otimes X_{\mathrm{B};l}$. Here $X_{\mathrm{A};l}$ denotes the operator composed by the previously defined X_{A} acting on qubits l and $l+1$ and the identity operator acting on the remaining qubits ($l = m$ means the first and last qubit). On Bob's side the operators $X_{\mathrm{B};l}$ are given by

$$X_{\mathrm{B};l} = \frac{1}{4}(|2l-1\rangle\langle 2l+1| + |2l+1\rangle\langle 2l-1| + |2l\rangle\langle 2l+2| + |2l+2\rangle\langle 2l|), \tag{3.51}$$

with addition being carried out modulo $2m$.

3.3.3(d) Communication channel model

In this part we present the employed channel model of the COW experiment used in our numerical simulations. Note, however, that the results presented here can be applied as well to any other quantum channel, as they only depend on the observed detection probabilities in both the data and monitoring lines.

In particular, we characterize the losses in the channel with a BS of transmittance η_{channel}. This parameter can be related with a transmission distance d measured in km for the given QKD scheme as

$$\eta_{\text{channel}} = 10^{-\frac{\alpha d}{10}}, \tag{3.52}$$

where α represents the loss coefficient of the channel (*e.g.*an optical fiber) measured in dB/km. Together with the efficiency of the detectors the overall system transmittance is given by

$$\eta_{\text{sys}} = \eta_{\text{channel}}\eta_{\text{det}}. \tag{3.53}$$

The total system loss in dB is used as the x-axis in the secret key rate figures, *i.e.*$-10\log_{10}\eta_{\text{sys}}$.

The channel misalignment is parametrized with an error probability e_{d} that a signal hits Bob's detectors in the wrong time slot within the same bit. For simplicity, we assume that e_{d} is a constant independent of the distance and we use $e_{\text{d}} = 1\%$ for simulation purposes. This effect is modeled by a completely positive trace-preserving map Φ that incoherently flips the signal states within the same bit slot as $\left|0, \sqrt{\eta_{\text{sys}}}\alpha\right\rangle \mapsto \left|\sqrt{\eta_{\text{sys}}}\alpha, 0\right\rangle$ and $\left|\sqrt{\eta_{\text{sys}}}\alpha, 0\right\rangle \mapsto \left|0, \sqrt{\eta_{\text{sys}}}\alpha\right\rangle$ with probability e_{d}. Here we consider that the input signals have been already affected by system losses. We have, therefore, that whenever Alice sends to Bob a corresponding COW signal state with coherent state $|\alpha\rangle$ in temporal mode d, the probability that Bob observes a single photon detection event in this mode only (within the whole signal block) is given by

$$p_d^{\text{correct}} = \text{tr}\{\Lambda_s[\Phi^{\otimes m}(\rho_{\text{loss}}^m)]M_d\} \tag{3.54}$$

$$= \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\text{sys}}\lambda} + (1-\epsilon)^{2m}(1-e_{\text{d}})\eta_{\text{sys}}\mu e^{-\eta_{\text{sys}}\lambda}, \tag{3.55}$$

where ρ_{loss}^m represents the output signal of the BS characterizing the total system loss, $\mu = \alpha^2$, and $\lambda = m\mu$. Similarly, when Alice sends to Bob a vacuum state in temporal mode d Bob can observe a single photon detection event in this mode only with probability

$$p_d^{\text{error}} = \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\text{sys}}\lambda} + (1-\epsilon)^{2m}e_{\text{d}}\eta_{\text{sys}}\mu e^{-\eta_{\text{sys}}\lambda}. \tag{3.56}$$

The total probability that Bob observes an inconclusive detection event in the data line is then given by

$$p_{\rm inc} = 1 - m\Big(p_d^{\rm correct} + p_d^{\rm error}\Big). \tag{3.57}$$

In the monitoring line we include an additional misalignment effect that reduces further the interferometric visibility. In particular, we assume that whenever two equal coherent states interfere at a $50:50$ BS then the outcome signal can exit the BS through the wrong output port with error probability $e_{\rm m}$. In our simulations we use $e_{\rm m} = 0.5\%$. Here we distinguish two possible scenarios, depending on whether the signals which interfere at the BS were prepared by Alice in the same quantum state or not. Let us assume that the first signal corresponds to bit i_l while the later to bit i_{l+1}. That is, Bob interferes modes $2l-1, 2l+1$ and, at the same time, $2l, 2l+2$.

Let us consider first the situation where both signals were generated in the same state $|0,\alpha\rangle$. In this scenario, we find that Bob observes a single photon detection event in temporal mode $2l-1$ only (and no click in the remaining modes of the block) with probability

$$\begin{aligned} p_{2l-1,+} &= \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\rm sys}\lambda} + (1-\epsilon)^{2m}\eta_{\rm sys}\mu e^{-\eta_{\rm sys}\lambda} \\ &\quad\times\Big[2(1-e_{\rm d})^2(1-e_{\rm m}) + e_{\rm d}(1-e_{\rm d})\Big], \\ p_{2l-1,-} &= \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\rm sys}\lambda} + (1-\epsilon)^{2m}\eta_{\rm sys}\mu e^{-\eta_{\rm sys}\lambda} \\ &\quad\times\Big[2(1-e_{\rm d})^2 e_{\rm m} + e_{\rm d}(1-e_{\rm d})\Big], \end{aligned} \tag{3.58}$$

where the superscript $\pm$ indicates whether the single excitation is registered in the bright (D_+) or in the dark (D_-) detector of the monitoring line. Similarly, we have that the probability that Bob sees a single photon detection in temporal mode $2l$ only is given by

$$\begin{aligned} p_{2l,+} &= \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\rm sys}\lambda} + (1-\epsilon)^{2m}\eta_{\rm sys}\mu e^{-\eta_{\rm sys}\lambda} \\ &\quad\times\Big[2e_{\rm d}^2(1-e_{\rm m}) + e_{\rm d}(1-e_{\rm d})\Big], \\ p_{2l,-} &= \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\rm sys}\lambda} + (1-\epsilon)^{2m}\eta_{\rm sys}\mu e^{-\eta_{\rm sys}\lambda} \\ &\quad\times\Big[2e_{\rm d}^2 e_{\rm m} + e_{\rm d}(1-e_{\rm d})\Big]. \end{aligned} \tag{3.59}$$

The case where both signals were generated in the same state $|\alpha,0\rangle$ is completely analogous. One only needs to interchange (3.58) and (3.59).

Finally, let us consider the situation where both signals are prepared in a different quantum state. In this scenario the probabilities are given by

$$p_{2l-1,+} = p_{2l,+} = \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\rm sys}\lambda} + (1-\epsilon)^{2m}\eta_{\rm sys}\mu e^{-\eta_{\rm sys}\lambda} \times \left[2e_{\rm d}(1-e_{\rm d})(1-e_{\rm m}) + \frac{1+2e_{\rm d}^2-2e_{\rm d}}{2}\right], \tag{3.60}$$

and

$$p_{2l-1,-} = p_{2l,-} = \epsilon(1-\epsilon)^{2m-1}e^{-\eta_{\rm sys}\lambda} + (1-\epsilon)^{2m}\eta_{\rm sys}\mu e^{-\eta_{\rm sys}\lambda} \times \left[2e_{\rm d}(1-e_{\rm d})e_{\rm m} + \frac{1+2e_{\rm d}^2-2e_{\rm d}}{2}\right]. \tag{3.61}$$

3.3.4 Simulation results

For simulation purposes, we consider that Bob's detectors are identical with a dark count rate of 10^{-7}. The channel model includes an intrinsic error rate of 1% in the data line together with an additional misalignment in the monitoring line that reduces the visibility to 99%. We study two different scenarios: (a) the case where all different m-signals blocks share the same phase, and (b) the scenario where each block is phase randomized. The resulting lower bounds on the secret key rate per pulse, $R_m/(2m)$, are illustrated in Figure 3.5. For comparison, this figure includes as well a lower bound on the secret key rate for a coherent-state version of the BB84 protocol with and without phase randomization [61, 62]. For a given total system loss, i.e. including the losses in the channel and in Bob's detection apparatus, we optimize the lower bound over the respective signal strength α of Alice's source which is of order 0.1. As expected, we find that case (b) performs better than that where all blocks share a common phase, since the signal states are less distinguishable for an eavesdropper without a global phase. We obtain that the tolerable system loss for the COW protocol is, respectively, ≈ 19.5 dB (a), and ≈ 22.6 dB (b). The bit error and visibility at these cutoff points are, respectively, $\approx 3\%$ and $\approx 96\%$ (a), and $\approx 5.3\%$ and $\approx 93.3\%$ (b).

Our simulations reveal that a main limiting factor in DPR-QKD seems to be the dark count rate of Bob's detectors. For given experimental parameters, there is an optimal finite block size that allows a maximum tolerable total system loss. If one increases the block size further this does not translate into an improved lower bound or distance. This is due to the fact that, in the high loss regime, large sized blocks suffer from a higher dark count probability *per block* than smaller sized blocks, and this reduces the achievable secret key rate. A similar effect was observed in the security analysis for the differential-phase-shift protocol with true

single photon sources [38]. For a dark count rate per pulse of 10^{-7} the optimal block size in the COW scheme turns out to be $m = 3$, *i.e.*6 optical pulses. Also, this figure shows that a coherent-state version of the BB84 protocol without decoy states can deliver higher key rates per signal than the analyzed COW protocol assuming the same channel. The reason for this might be threefold: (1) the small optimal block size in the COW scheme; (2) considering blocks, it can be shown that certain multi-photon pulses are completely insecure; (3) most importantly, while in the BB84 the phase error is measured directly, in the COW protocol it has to be estimated.

3.3.5 Conclusions

We have presented a generic method to prove security of practical DPR-QKD against general attacks. With the explicit example of a variant of the COW protocol, we have shown that these schemes are indeed secure for certain distances at given rates. Its performance, however, seems to be less robust against practical imperfections than originally expected.

To further improve the lower bounds shown in Figure 3.5 there are several alternatives. Since a main limitation seems to come from dark counts, one may consider security in the fully calibrated device scenario where these errors are not attributed to Eve. As a quantitative bound on the performance of this scenario we investigated the case of a zero dark count rate, in which all bounds shown in Figure 3.5 shift by about 3 dB, though the difference between the COW and the BB84 protocol remains. Additionally, one can evaluate different announcements in a similar spirit like the SARG protocol [63]. We considered different declarations, but unfortunately none of them enhanced the resulting rate. Another possibility is to include, for instance, an extra monitoring line on Bob's side to additionally check the coherence between subsequent pulses. The state distribution part of this protocol is then very similar to the original COW scheme with an additional decoy signal composed by two vacuum pulses [64]. This hardware change improves the maximum tolerable system loss by about 1 dB.

Another hardware change might be to include additional phase differences in the signal stream, such that the signals states get closer to the one used in a BB84 protocol. Finally, one may ask whether different security techniques might provide better lower bounds. For instance, one could consider more valid detection events per block. This needs however much larger block sizes such that one obtains at all a reasonable fraction of two or more click events in the long distance limit. Another alternative would be to bound the rate by the individual phase errors, i.e. directly

using (3.22). This could give a benefit if, for example, bits at the boundary are much easier to infer by Eve than bit values originating from events well inside the block. Clearly another option would be to abandon the block idea. However even in this case Eve could always attack the signals block-wise. Though a coherence measurement across blocks would then reveal the eavesdropper, any coherence measurement within would be still fine. Hence considering only an average visibility this effect will become less and less important. All these alternatives definitely deserve further investigations, but we do not expect a dramatic improvement.

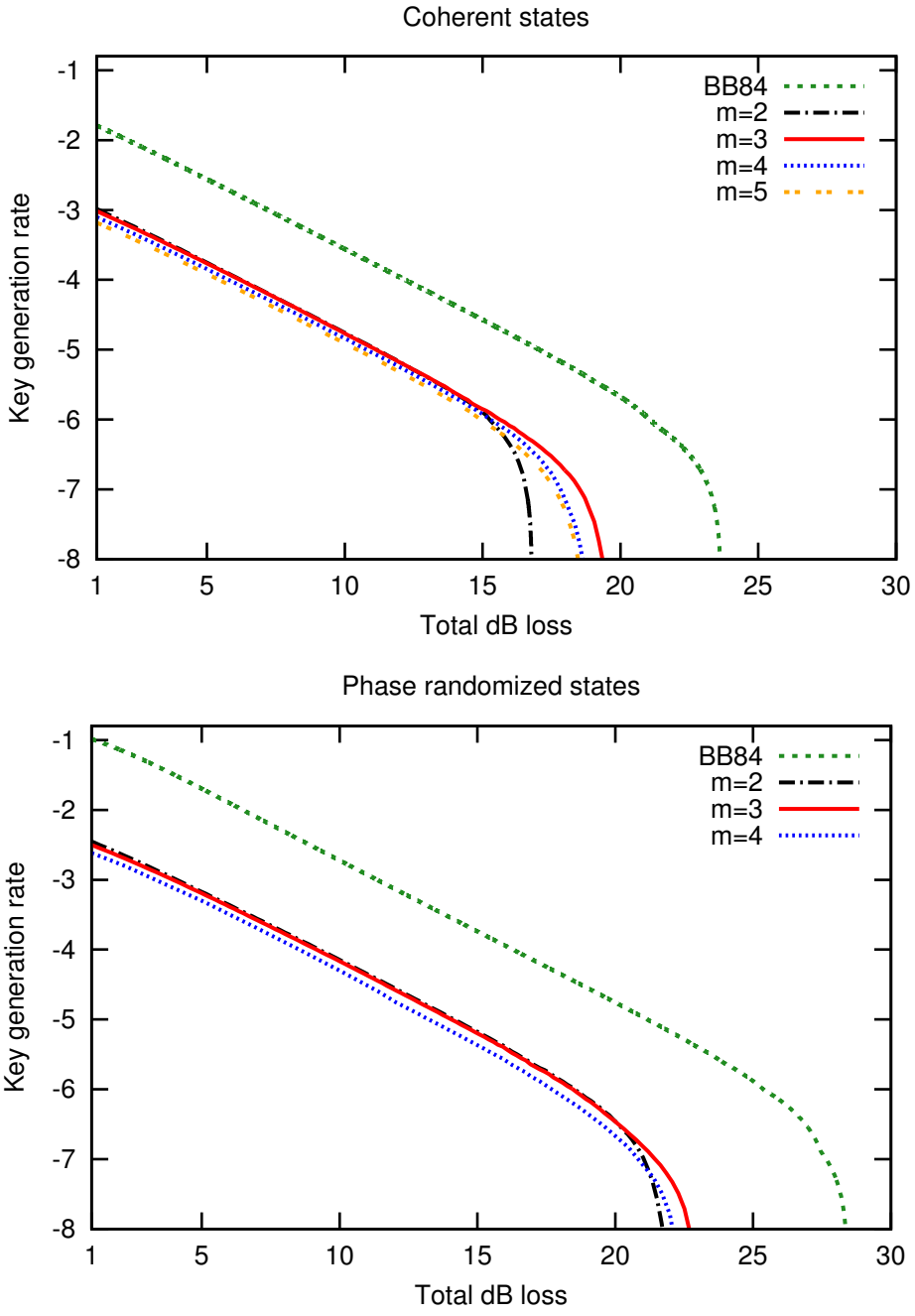

Figure 3.5: Lower bound on the secret key rate given by (3.23) per pulse on a logarithmic scale (base 10) vs. the total system loss in dB for the COW protocol illustrated in Figure 3.2 using signal blocks carrying m bits of information (i.e. $2m$ optical pulses) in the security proof. The upper figure corresponds to the case where all blocks of signals share a common phase, while in the lower figure each block is phase randomized. For comparison, we include a lower bound on the secret key rate for a coherent-state version of the BB84 protocol [3] with and without phase randomization [61, 62]. We consider three main errors: an intrinsic error rate of 1% in the data line, an additional misalignment in the monitoring line reducing the visibility to 99%, and a dark count rate of 10^{-7} for each detector. Moreover, in the lower figure we assume $n_{\text{cut}} = 2$.

CHAPTER 4

QUANTUM RANDOMNESS GENERATION

4.1 Randomness from different levels of characterization

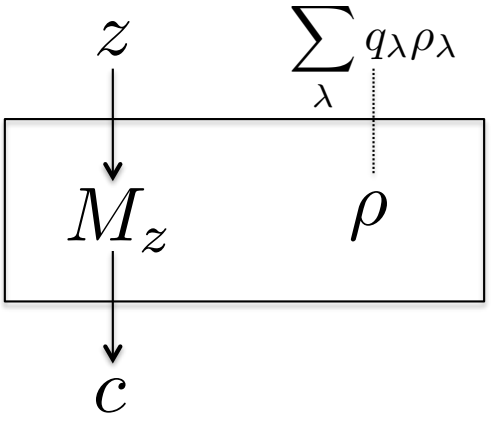

Figure 4.1: Generic setup of quantum randomness extraction. The authorized party inputs z and receives the outcome c, whose randomness one wants to guarantee with respect to an adversary Eve. Inside the box, the role of z consists in selecting a possible measurement M_z to be performed on a state ρ, and c is going to be the outcome of that measurement. Various scenarios can be considered based on the power given to Eve and the level of characterization that the authorized party has of her devices.

The generic setup for quantum randomness is sketched in Figure 4.1. As explained in the caption, the goal of the authorized party is to certify the randomness of the outcome c *with respect to Eve's knowledge*. There is a priori place for a third party, the provider, who may have produced some of the devices but has no

interest in learning c. A scenario will be defined by the power given to Eve and the level of characterization of the devices. We address them in this order.

For definiteness, we focus exclusively on the case of *measurement independence*: namely, we assume that the state to be measured ρ and the choice of the measurement z are fully uncorrelated in each run. It is remarkable that even this assumption can be partially relaxed, giving rise to the possibility of randomness amplification [32, 33, 39, 65–69]. However, our results in Section 4.2 show that any Bell-based randomness amplification protocol cannot amplify arbitrary input source.

4.1.1 Scenarios for quantum randomness

4.1.1(a) Classes of adversarial power

Throughout this section, we are going to assume that quantum theory holds; in particular, we don't discuss the possibility of certifying randomness against an adversary limited only by no-signaling [30, 32, 33, 66, 67, 70]. The power given to Eve can be divided in three main classes:

Class (I) Eve is outside of Alice's laboratory. Because of the Kerchhoff-Shannon principle (one should not look for security in hiding details of the hardware or the protocol), we assume *Eve has knowledge of the experimental setup but cannot influence it.* Eve may know in each run which quantum state enters the box, and which measurement is used (not only the set of possible measurements). Her description of both state and measurements may be better than that of both Alice and the provider: for instance, she may be able to describe the state as pure in each run by knowing which decomposition of a possible mixture is being used. Since the choice of measurements in each run is known to Eve, we speak of *randomness generation.* We can define two subclasses (a) and (b), according to whether Eve does not (classical side information) or does hold (quantum side information) a possible purification of the state in a quantum memory (in the literature on quantum cryptography one finds also intermediate situations between (a) and (b), the so-called bounded-storage models [71] and noisy storage models [72]). Because quantum theory is no-signaling, holding a purification does not allow Eve to change the state ρ, but may give her more guessing power since she will be able to steer towards a specific decomposition (by performing a specific measurement on her reduced state).

Class (II) Eve is a party inside Alice's laboratory. In other words, we can allow *Eve to distribute the state*, while giving her only classical, though perfect, knowledge about the set of possible measurements. If Eve is allowed to know the choice of measurement in each run, she could use her power to distribute a state that gives her a deterministic outcome (and measurement independence is violated). In other words, randomness generation is impossible for an adversary in this class. To be compatible with measurement independence, Eve is not allowed to know which measurement is being used in each run, and we speak of *randomness expansion* where an initial seed of randomness with respect to Eve must be required. We can again define subclasses (a) and (b) as above. Class II(b) is a natural analog of the entanglement-based scheme in which quantum cryptography can be proven "unconditional security": as such, most papers on quantum randomness have considered it [29–31, 73]. This class is natural in the case where Alice holds two subsystems situated in secured, but possibly distant locations.

Class (III) The maximal class is that in which *Eve is the provider*, namely she prepares both the measurement devices and the state: the only power left with Alice is the choice of z in each run. The claim is frequently made, at least in semi-popular accounts, that quantum physics could provide security against an adversarial provider. The correct statement is that one may be able to *certify the quantumness* of the process that generates c, typically in the case of a loophole-free Bell test. However, an adversarial provider would certainly find a way to *hide a transmitter* in the devices, with the task of leaking out the values of c at the end of the protocol, or may employ attacks discussed in [74]. Therefore, however certifiably quantum the process that generated the outcomes may have been, there cannot be any randomness with respect to an adversarial provider. So we won't consider this class any longer in this thesis.

A further point worth reiterating is the amount of randomness needed to generate the seed z. What must be guaranteed is measurement independence, i.e. the choice of z must be "random" with respect to the choice of the state in each run, and vice versa; while the randomness we want to extract must be unpredictable for Eve. Therefore, if one works in Class I(a), no randomness with respect to Eve is required in the seed: even if Eve knows z, she cannot adapt the state to it. One can speak of *randomness generation* [75]. On the contrary, in Classes I(b) and II, since Eve steers or prepares the state, the inputs z must be generated with a process that is *a priori* guaranteed to be unknown to Eve. In other words, in these

Eve...	Class I *has no access to the devices*	Class II *distributes the state*	Class III *prepares the devices*
Subclass (a) *holds classical side information*	randomness generation	randomness expansion	no randomness
Subclass (b) *holds quantum side information*	randomness generation		

Table 4.1: Randomness protocols in presence of measurement independence with different classes of adversaries. The gray cell corresponds to the trusted provider assumption, which describes the class of adversaries considered in this work.

classes there is no randomness generation, but only *randomness expansion* — and, after a series of partial results [29, 75, 76], it has been proved that such expansion can in principle be unbounded [30].

Finally recall that we are assuming strict measurement independence. If partial measurement dependence is taken into account, one would have to refine the definition of these classes: for instance, by specifying whether the dependence is introduced unwittingly by the provider or maliciously by Eve.

4.1.1(b) Levels of characterization of the devices

The second defining feature of a scenario is the level of characterization that Alice has of the working of her devices. We sketched it in the introduction, now we can be more precise:

- The *tomography* level of characterization usually means that Alice knows the behavior of her devices as well as Eve does. This is what we shall consider in this paper. Of course, in the vast literature on quantum tomography, the possibility of partial knowledge of the devices has been considered [77–80], so one could refine this level of characterization in several sublevels. In all cases, though, it is assumed that Alice knows exactly which degrees of freedom are relevant to the measurement (polarization, spin, quadrature of a field...).

  When this trust in the characterization is unwarranted, a Class II Eve may successfully *hack* the devices. In particular, Eve may know that some degrees of freedom, others than the ones Alice is aware of, play a role in the physical process, and may thus influence the behavior of the boxes by addressing those degrees of freedom. For instance, this was the case in the series of experiments that hacked quantum cryptography devices by exploiting the physics of photodetectors [81–83]. Similarly, a Class I Eve could also take advantage of her knowledge about what happens in additional degrees of freedom if a setup happens to use them.

- The *device-independent* level of characterization means that Alice does not rely on a description of her devices but only on the observed statistics. Since the statistics on a single quantum system can always be reproduced with classical randomness, device-independence requires a loophole-free violation of a Bell inequality. Specifically, it is crucial to close the detection loophole, while the no-signaling condition may be justified in other ways than by arranging spacelike separation. In any case, Alice may actually hold both measurement devices in her lab. Nevertheless, for convenience we shall speak of Alice and Bob when it comes to device-independence. We also note that, as it happens for measurement independence, the strict no-signaling condition may be partially relaxed, but we don't consider this relaxation in this paper [84].

- One can define intermediate levels of characterization, collectively known as *semi-device-independent*. For instance, in the context of quantum cryptography, the idea of *measurement-device-independence* has been put forward after noting that hacking usually involves detectors (which are therefore better left untrusted) rather than sources (which may therefore be trusted) [26]. Other works relax the tomographic requirement of perfect knowledge of the degree of freedom, but assume an upper bound on the dimensionality of the systems under study [85]. In this paper, we shall consider explicitly the case called *one-sided device-independent*, in which entanglement is certified by two devices, Alice's being tomographic and Bob's unknown; this case was studied for quantum key distribution [86].

4.1.1(c) Assumptions of this Section

In this paper, we consider Eve in *Class I(a)*, which was called "trusted provider" in previous works [87]. In real life, this class describes randomness generation in an experiment performed by a trusted laboratory: therefore, even if it does not explore the ultimate limits of quantum power, it is arguably relevant for physics [88, 89]. The randomness is guaranteed against any Eve that is not involved in setting up or running the experiment. Furthermore, Eve not being in the lab, she could hold the purification only if those degrees of freedom were "radiative" and she had the power to collect them: this, together with the state-of-the-art of quantum memories, makes it very reasonable to restrict to subclass (a) at least for a few years to come.

Also, we consider only the asymptotic limit of infinitely many runs where each run is assumed to be independent and identically distributed (i.i.d.) according to

some strategy of Eve. Corrections due to finite samples, and extension to non-i.i.d. scenarios can in principle be done with the techniques in [75, 90, 91].

4.1.2 Computing randomness for different levels

4.1.2(a) Definitions and notations

Let us introduce the basic notions to study randomness (cf. [92]). The authorized party can input $z \in \{1, ..., m\}$ in the box and obtain output $c \in \{1, ..., d\}$ (see Figure 4.1). In the asymptotic limit of infinitely many runs, she can reconstruct the statistics $P(c|z)$. This statistical distribution may reflect either accidental randomness (due to ignorance of some details of the state or the device) or intrinsic randomness, due to the unpredictability of the outcome of quantum measurements. We are interested in the latter because for Eve, in any of the Classes defined above, there is no accidental randomness.

In this section, the states $|\psi\rangle$ or ρ (single- or bi- or multi-partite) refer to everything that is in the box, so that the measurements can be assumed to be projective. When the box is fully characterized, it becomes possible to distinguish between the system and a possible ancilla used for POVM measurements. We analyze the randomness of this case in 4.1.4(b)

If the state is pure, there is no accidental randomness for von Neumann measurements. Then, the randomness of c obtained from a given z is quantified by the probability $G(|\psi\rangle, z)$ of guessing the outcome correctly. Since the best strategy for guessing is to guess the most probable outcome, this guessing probability is

$$G(|\psi\rangle, z) = \max_c P(c|z, |\psi\rangle), \tag{4.1}$$

where $P(c|z, |\psi\rangle)$ is the probability of obtaining the outcome c given a measurement z on quantum state $|\psi\rangle$. If the state shared by Alice and Bob is mixed, we have to separate the intrinsic randomness from the accidental one. For measurement z, the average guessing probability that quantifies intrinsic randomness is then given by

$$G(\rho, z) = \max_{\{q_\lambda, \psi_\lambda\}} \sum_\lambda q_\lambda G(|\psi_\lambda\rangle, z), \tag{4.2}$$

where $\rho = \sum_\lambda q_\lambda |\psi_\lambda\rangle \langle\psi_\lambda|$, and the maximization is taken over all possible such decompositions.

For the device-independent level of characterization, Alice cannot write down a quantum state but needs to use only the observed probability distribution $P(c|z)$.

Then the guessing probability that quantifies intrinsic randomness is

$$G(P, z) = \max_{(\rho,M)\to P} G(\rho, z), \tag{4.3}$$

where the maximization is taken over all quantum states ρ and measurements M compatible with the probability distribution P.

In all these cases, the number of random bits that can be extracted per run is quantified by the min-entropy (c.f. the preliminary chapter)

$$H_{\min}(G) = -\log_2 G\,. \tag{4.4}$$

Finally, one may consider extracting randomness out of several settings, rather than a single one. If Eve is allowed to keep a purification of the state [subclasses (b)], upon learning which settings have been used in a given run, she can steer the state to the decomposition that maximizes (4.2) for those settings. In this case, therefore, there is no advantage for Alice to use several settings: she should just stick to the setting z that gives the smallest G. If Eve does not hold a purification [subclasses (a)], however, using several settings is advantageous [87].

Now we are going to explain how randomness can be computed in each of the three levels of characterization of our concern. For the device-independent level of characterization, randomness generation against a Class Ia Eve have been presented in [87, 93], so we don't repeat this here.

4.1.2(b) Tomography level of characterization

For the case of tomography level of characterization, we are on a ground familiar for most physicists. If a qubit is prepared in the state $|+z\rangle$, to say that a measurement of σ_x provides a perfect random bit is just a rephrasing of elementary textbook knowledge. The example of a qubit prepared in the maximally mixed state $\mathbb{1}/2$ is only slightly more involved: then, a measurement of a single observable (say) σ_x does not guarantee any randomness, because the state may have been prepared by mixing eigenstates of that operator, in which case Eve would have full knowledge of the outcome of each run. However, the uncertainty relations provide a way around it: if in each run Alice can choose to measure either σ_x or σ_z, no preparation can be an eigenstate of both, therefore there is randomness with respect to Eve as long as she does not hold a purification (see [94] for how uncertainty relations must be modified if Eve does hold a purification).

Now we provide a general recipe to compute the intrinsic randomness for projective measurements; the case of POVMs will be discussed in 4.1.4(b).

Since the state ρ can be reconstructed and is therefore part of the observed data, we need to perform the maximization of (4.2). In the decomposition, it is not a priori obvious how many quantum states $|\psi_\lambda\rangle$ are to be considered. Fortunately, the argument used in [87] can be transposed directly from probability distributions to density matrices, so it is sufficient to consider *one state per outcome*. Explicitly, for a projective measurement $M \equiv \{\Pi_c, c = 1, ..., d\}$ with d outcomes, the optimal guessing probability reads

$$G(\rho, M) = \max_{\{q_\lambda, \psi_\lambda\}} \left[\sum_\lambda q_\lambda \max_c \mathrm{tr}(|\psi_\lambda\rangle \langle\psi_\lambda| \Pi_c) \right]. \tag{4.5}$$

Define the subnormalized state $\rho_c = \sum_{\lambda \in \Lambda_c} q_\lambda |\psi_\lambda\rangle \langle\psi_\lambda|$ where $\Lambda_c = \{\lambda \in \Lambda : \arg\max_{c'} \mathrm{tr}(|\psi_\lambda\rangle \langle\psi_\lambda| \Pi_{c'}) = c\}$ forms a partition for Λ, then

$$G(\rho, M) = \max_{\{\rho_c\}} \sum_c \mathrm{tr}\left(\rho_c \Pi_c\right) \tag{4.6}$$

under the constraints that $\rho = \sum_c \rho_c$, $\rho_c \geq 0$. Like in the case of device-independence, this maximization is a semi-definite program (SDP), the only difference being that the matrix that must be positive is the quantum state itself, not a moment matrix of the observed statistics. Moreover, here the SDP solves the problem of interest directly, rather than a relaxation thereof. Given the tomography level of characterization, Alice can choose to extract randomness from *any* measurement, and will choose that for which the guessing probability in (4.6) is the lowest. Hence, for a given state ρ the optimal randomness extractable from measuring a single von Neumann measurement on this state is

$$G(\rho) = \min_M G(\rho, M) \text{ [one measurement]} \tag{4.7}$$

with the minimization is over the set of possible projective measurements on the system. Further, when Eve is not allowed to hold a purification, it may be advantageous to extract randomness from more measurements. If measurement M_z is chosen with probability q_z, the average guessing probability will be

$$G(\rho, \{M_z\}_z) = \max_{\{\rho_C\}} \sum_C \sum_{z=1}^m q_z \,\mathrm{tr}\left(\rho_C \Pi^z_{c_z}\right) = \max_{\{\rho_C\}} \sum_C \mathrm{tr}\left(\rho_C \mathcal{M}_C\right), \tag{4.8}$$

where we have denoted $\mathcal{M}_C = \sum_{z=1}^m q_z \Pi^z_{c_z}$; the constraints are as above, and now $C = (c_1, c_2, ..., c_m)$, so the maximization now involves a decomposition on d^m states. The fact that Eve cannot steer Alice's mixture is explicit in that the decomposition $\rho = \sum_C \rho_C$ is independent of z. As above, Alice is allowed to choose the set of

measurements that minimizes the guessing probability, so

$$G(\rho) = \min_{\{M_z\}} G(\rho, \{M_z\}) \text{ [more measurements]}. \tag{4.9}$$

Notice that unlike the optimization (4.8) which is an SDP, this last optimization over the choices of measurement settings does not appear to be an SDP.

4.1.2(c) One-sided device-independent level of characterization

The one-sided device-independent level of characterization, to our knowledge, has never been considered before in the context of randomness. The scenario is very similar to steering [95, 96]: the setup is actually the same, but the figure of merit is different. Indeed, instead of having Alice to convince Bob that she can steer his state, we just let them perform their measurements locally and ask whether randomness can be extracted from their outcomes.

Like before, we consider first the amount of random bits that can be extracted from the outcomes $c = (a, b)$ of a single pair of measurements $z = (x, y)$ with $A_x = \{\Pi_a^x\}$ and $B_y = \{\Pi_b^y\}$ denoting Alice and Bob's local measurements. The guessing probability is analog to (4.3) and given by

$$G(P, z) = \max_{(\rho, A_x, B_y) \to P} G(\rho, A_x, B_y) = \max_{(\rho_c, A_x, B_y) \to P} \sum_c P_c(c|z) \tag{4.10}$$

where $\rho = \sum_c \rho_c$, and $P_c(a, b|x, y) = \text{tr}(\rho_c \Pi_a^x \otimes \Pi_b^y)$. The constraints for the optimization are the observed statistics $P(a, b|x, y)$, and the knowledge of the state and measurements on Bob's side.

Such optimization is very similar to the one used for the device-independent level of characterization in [87, 93], where one can use the hierarchy introduced in [49] to provide upper bounds. In that case, from the set of local measurements and depending on the hierarchy's level, one forms a certain matrix Γ_c whose elements are expectation values with ρ_c of products of operators of the form: $\langle M_A \otimes \mathbb{1} \rangle$, $\langle \mathbb{1} \otimes M_B \rangle$, $\langle M_A \otimes M_B \rangle$, $\langle M_A M'_A \otimes \mathbb{1} \rangle$, $\langle \mathbb{1} \otimes M_B M'_B \rangle$, $\langle M_A M'_A \otimes M_B M'_B \rangle$, etc, where M_A, M'_A, and M_B, M'_B are operators from the set of Alice's and Bob's local measurements (union the identity), respectively (see [49] for a detailed description of this matrix). Some elements of Γ_c are related to the $P_c(a, b|x, y)$ mentioned above, while others are extra unknown variables in the optimization. By constraining Γ_c to be positive semi-definite and the sum over c of $P_c(a, b|x, y)$ to be the observed statistics, one can bound the guessing probability in the device-independent level of characterization.

Now in the one-sided device-independent case, we impose further constraints on the elements of Γ_c based on the knowledge of Bob's measurements. Namely, we use

the algebraic relations satisfied by these operators to constraint the moments of Γ_c which involve them. For instance, if $B_3 = (B_1 + B_2)/\sqrt{2}$ or $B_1B_2 = -B_2B_1$, the relations $\langle \mathcal{O}B_3\rangle = (\langle \mathcal{O}B_1\rangle + \langle \mathcal{O}B_2\rangle)/\sqrt{2}$ or $\langle \mathcal{O}B_1B_2\rangle = -\langle \mathcal{O}B_2B_1\rangle$ are imposed for all product of operators $\mathcal{O}$. This reduces the number of independent variables in the optimization. These relations are imposed on each c in (4.10), as they should hold for each of the ρ_c in the decomposition. Note that we do not directly use the knowledge of Bob's local state to further constraint the optimization. We thus obtain an upper bound on the guessing probability.

Just like in the other levels of characterization, one could consider extracting randomness from more than one measurement here, if Eve does not hold a purification of the measured state. The optimization problem can be set up in a manner analogous to what we have been considering.

4.1.3 Comparison of the yields of three levels

In order to compare the yields of the various levels of characterization, we need a common set of data. We assume that the data come from measuring a two-qubit Werner state $\rho_V = V\left|\Phi^+\right\rangle\left\langle\Phi^+\right| + (1-V)\mathbb{1}/4$; Alice measures either $A_1 = \sigma_x$ or $A_2 = \sigma_z$, Bob measures one of the four $B_1 = \sigma_x$, $B_2 = \sigma_z$, $B_3 = \sigma_+$ or $B_4 = \sigma_-$ with $\sigma_\pm = (\sigma_x \pm \sigma_z)/\sqrt{2}$. These measurements can lead to non-trivial assessment for all the level of characterization we are interested in. Indeed, $(A_1, A_2; B_1, B_2)$ can be used for partial tomography and identify $\left|\Phi^+\right\rangle$ uniquely when $V = 1$; for device-independence, $(A_1, A_2; B_3, B_4)$ violate the CHSH inequality for $V > 1/\sqrt{2}$ and certify $\left|\Phi^+\right\rangle$ for $V = 1$ because CHSH $= 2\sqrt{2}$ [97]; for one-sided device-independence, a similar argument holds for $(A_1, A_2; B_1, B_2)$ and the steering inequality S_2 defined in [98]. Anyway, in what follows, the amount of randomness is computed directly from the observed statistics, without processing them into a specific tomography protocol or inequality.

As mentioned before, we focus on the randomness generated by a single pair of setting. In the device-independent case, we bound the amount of random bits using the second level of the hierarchy [49], as described in (4.3) (see [87, 93] for more details). For the one-sided device-independent case, we use the method described in 4.1.2(c) at the same level of the hierarchy, together with the algebraic relations generated by Bob's measurements (including $B_3 = (B_1 + B_2)/\sqrt{2}$, $B_4 = (B_1 - B_2)/\sqrt{2}$, $B_1B_2 = -B_2B_1$, $B_3B_4 = -B_4B_3$, etc.). Finally, for tomography characterization, we compute the amount of randomness based on (4.8), as explained in 4.1.2(b).

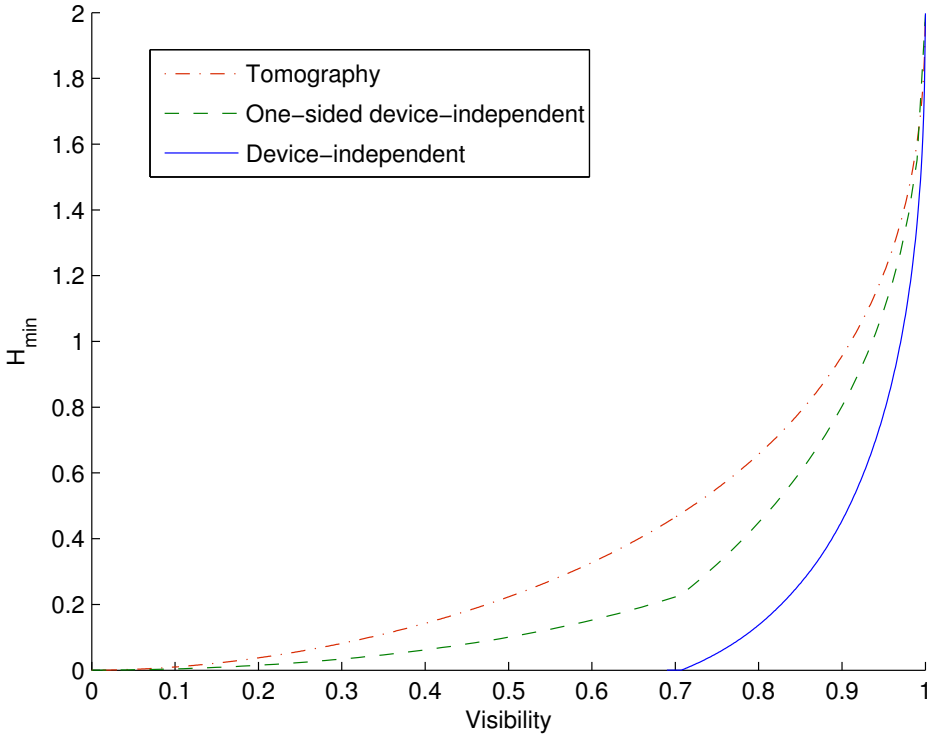

Figure 4.2: Amount of randomness extracted $H_{\min}$ from the outcomes of the setting pair A_2, B_1 from three different levels of characterizations.

The main result we obtained for three different levels of characterization is shown in Figure 4.2. As expected, for any visibility V the amount of randomness increases with the level of characterization of the devices, with the tomographic level giving the largest amount of randomness.

Note that for all three different levels of characterization, two bits of randomness can be extracted when $V = 1$. In the device-independent case, this is more randomness than the ~ 1.23 bits that can be certified from the maximal violation of the CHSH inequality [28]. To understand this difference, we compare the amount of randomness that can be certified in this scenario from different constraints.

Namely, we consider the randomness that can be certified from the correlations above, the one that is certified from an optimal violation of the CHSH inequality with the same state, and from an optimal violation of a modified CHSH inequality

$$\mathrm{CHSH}_3 = \langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle + \langle A_1B_3\rangle \leq 3\,. \tag{4.11}$$

These last two computations are performed by fixing the value of the inequality rather than the value of the correlations in the corresponding SDP. The result is shown in Figure 4.3: 2 bits of randomness can be extracted indeed from CHSH_3

when the pair of perfectly uncorrelated measurements Z_A, X_B are used. However, no such measurements are available when CHSH is maximally violated.

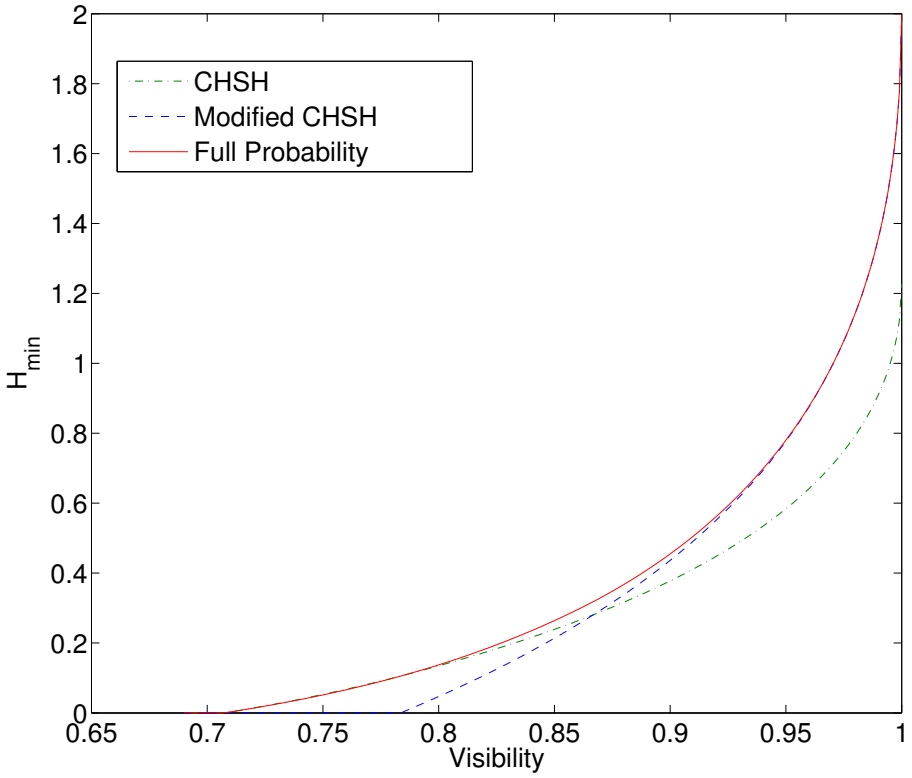

Figure 4.3: The amount of randomness computed with different constraints: CHSH, CHSH_3, and full observed statistics. $H_{\min}$ of CHSH corresponds to the setting pair A_2, B_3, while others correspond to A_2, B_1.

Note also that in Figures 4.2 and 4.3, randomness can be extracted in the device-independent case only provided that a Bell inequality is violated, i.e. when $V > 1/\sqrt{2}$ for CHSH and $V > 3/(2\sqrt{2}+1) \approx 0.78$ for CHSH_3 [1]. In other words, no randomness is found when a local hidden variable model can produce the observed correlations.

In the one-sided device-independent case, however, randomness can be extracted from all Werner state, except the completely mixed state. Yet, it is known that Werner states are non-steerable for $V \leq 0.5$ [95], i.e. such states admit a Local Hidden State (LHS) model. Thus, our result shows that one can certify randomness in one-sided device-independent context even in presence of a LHS model.

This can be understood by the fact that a local hidden state model only ascribes fixed outcomes to the measurement of one party. The other one, Bob in our case,

[1] For Wehner state, the value of CHSH_3 as a function of the visibility is $V(2\sqrt{2}+1)+(1-V)0$.

receives a quantum state to measure. However, Bob enjoys in this context a tomographic level of characterization of his system. He can thus always extract some randomness from this state. This is the case unless all the quantum states given to Bob can all be chosen in the same basis, as possible when $V = 0$.

Note that randomness can be extracted for all $V > 0$ in the tomographic level of trust as well. However the randomness also disappears there when $V = 0$, i.e. when the Werner state is white noise. In the next section, we discuss how randomness can be extracted from the white noise by using several measurement settings.

4.1.4 More results on the tomographic level

4.1.4(a) Randomness from single-qubit white noise and uncertainty relations

Among the many possible illustrations of randomness in the tomographic level of characterization, we consider the case of randomness extraction from a single qubit in the maximally mixed state $\rho = \mathbb{1}/2$, against an adversary of Class I(a). We recover known results on uncertainty relations and show numerical evidence for more general situations.

As mentioned at the beginning of 4.1.2(b), for a single measurement (equation (4.7)) one has $G(\mathbb{1}/2) = 0$: for whatever measurement being performed, the mixture may have been prepared by mixing the two eigenstates of the measurement and this information could be available to Eve. So to bound the guessing capability of Eve, we need to consider *more than one measurements*. We thus refer to (4.8) and (4.9) from now onwards.

Let us denote $\{M_k\}_{k=1,...,N}$ the N projective measurements with outcomes ± 1 on a qubit which is characterized by their Bloch vector $\vec{n}_k$. For any string of values $C = (c_1, ..., c_N) \in \{-1, +1\}^N$, the effective measurement operator is

$$\mathcal{M}_C = \frac{\mathbb{1} + \vec{n}_C \cdot \vec{\sigma}}{2} \quad \text{with} \quad \vec{n}_C = \sum_{k=1}^{N} q_k \, c_k \, \vec{n}_k \ . \tag{4.12}$$

With this notation, we have

$$G(\mathbb{1}/2, \{M_k\}) = \frac{1 + \max_C |\vec{n}_C|}{2} \ . \tag{4.13}$$

Indeed, the r.h.s. is obviously an upper bound, since it is the largest of the eigenvalues; and it can be achieved by the decomposition $\mathbb{1}/2 = \frac{1}{2} |+\vec{n}_{\bar{C}}\rangle \langle +\vec{n}_{\bar{C}}| + \frac{1}{2} |-\vec{n}_{\bar{C}}\rangle \langle -\vec{n}_{\bar{C}}|$ where $\bar{C}$ is defined by $|\vec{n}_{\bar{C}}| = \max_C |\vec{n}_C|$. Finally, we are allowed to choose the N most favorable measurements, i.e. (4.9) becomes $G(\mathbb{1}/2, N) = \frac{1+g_N}{2}$

for

$$g_N \equiv \min_{\{M_k,q_k\}} \max_C |\vec{n}_C| \,. \tag{4.14}$$

Now, since $|\vec{n}_k| = 1$, we have

$$|\vec{n}_C|^2 = \sum_k q_k^2 + \sum_{k\neq k'} (c_k q_k \vec{n}_k) \cdot (c_{k'} q_{k'} \vec{n}_{k'}) \,. \tag{4.15}$$

Notice that the second term can always be made non-negative by the maximization over C. Indeed, it follows from

$$\sum_{c_1,\ldots,c_N} \sum_{k\neq k'} (c_k q_k \vec{n}_k) \cdot (c_{k'} q_{k'} \vec{n}_{k'}) = 0 \tag{4.16}$$

that $\max_C \sum_{k\neq k'} (c_k q_k \vec{n}_k) \cdot (c_{k'} q_{k'} \vec{n}_{k'}) \geq 0$. Therefore, in the minimization, the best choice would consist in choosing all the vectors mutually orthogonal, but this is possible only for $N = 2, 3$. In these cases, it is simple to finish the optimization: we find $g_2 = 1/\sqrt{2} \approx 0.7071$ and $g_3 = 1/\sqrt{3} \approx 0.5774$. Notice that, translated in min-entropy, the case for $N = 2$ bound saturates the uncertainty relation for two min-entropies, equation (9) of [99], namely $H_{\min}(\sigma_z) + H_{\min}(\sigma_x) \geq \log(\frac{1+1/\sqrt{2}}{2})$.

To go further, we resort to numerical optimization to obtain upper bounds on g_N. For $N = 4$, the optimal choice of measurements is found to be $\{(\sigma_z, 1-3q), (\sigma_1, q), (\sigma_2, q), (-\sigma_3, q)\}$ where the vectors $\vec{n}_{1,2,3}$ are 120 degrees apart from each other in the $x-y$ plane; knowing this geometry, one can finish the optimization analytically to find $g_4 = \sqrt{4/13} \approx 0.5547$ for $q = 3/13$. For $N = 5$ and $N = 6$, we find $g_5 \approx 0.5422$ and $g_6 \approx 0.5270$. This trend suggests that, when $N \to \infty$, one has $g_N \to \frac{1}{2}$. This would be the case if the optimal choice would consist in spreading the $\vec{n}_k$ uniformly in the half-sphere [2]. Indeed, parametrizing the measurement Bloch vector using polar coordinates $\vec{n}(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ and using the Haar measure on half of the sphere $\mathrm{d}\theta\mathrm{d}\phi \sin\theta/2\pi$ we have

$$\vec{n}_C = \int_0^{\pi/2} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \frac{\sin\theta}{2\pi} (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) = (0, 0, 1/2). \tag{4.17}$$

4.1.4(b) Randomness from POVMs

As we have just seen and also mentioned in 4.1.2(b), no randomness can be extracted from a single projective measurement on the maximally mixed state, because Eve may know the decomposition in the eigenvalues of that measurement.

[2] This guess comes from the intuition that for all measurement configuration $\{M_k, q_k\}$ the inner maximization over C occur at $c_1 = c_2 = \ldots = c_N = +1$.

The reasoning does not seem to apply to POVMs, though: even knowing a pure state, in general Eve cannot guess with certainty the outcome of a non-projective POVM (for instance, the state $|\psi\rangle$ and POVM $M = \{\mathbb{1}/d, \mathbb{1}/d, ..., \mathbb{1}/d\}$). Is it therefore possible to extract randomness from a single POVM on the maximally mixed state? The answer is, yes, but the origin of the randomness makes the problem trivial. Indeed, because of Neumark's theorem, a POVM is nothing else than a projective measurement on the system and some additional degrees of freedom. We are going to show that a POVM on the maximally mixed state cannot provide more randomness than that present in the ancilla — thence it is pointless to perform the POVM for randomness purposes, one could have measured the ancilla directly.

To see this, let us first denote the system's dimensions by d_s. The n-outcomes POVM elemenets are written as $\{\Pi_c\}_{c=1,...,n}$. Given that the state we have is white noise, the probability for Eve to guess the outcomes correctly is

$$G(\mathbb{1}/d_s, \{\Pi_c\}) = \max_{\{q_c, |\psi_c\rangle\}} \sum_{c=1}^{n} q_c \operatorname{tr}(\rho_c \Pi_c) \tag{4.18}$$

where $\mathbb{1}/d_s = \sum_c q_c \rho_c$ and, as previously, ρ_c groups all the states in the decomposition for which $\max_{c'} \operatorname{tr}[\rho \Pi_{c'}] = \operatorname{tr}[\rho \Pi_c]$. Since the maximization in (4.18) is taken over all possible decomposition, any specific decomposition provides a lower bound. In particular, we can choose

$$\rho_c = \frac{\Pi_c}{\operatorname{tr}(\Pi_c)}, \quad q_c = \frac{\operatorname{tr}(\Pi_c)}{d_s} \tag{4.19}$$

which indeed gives $\sum_c q_c \rho_c = \mathbb{1}/d_s$. Then

$$G(\mathbb{1}/d_s, \{\Pi_c\}) \geq \sum_{c=1}^{n} \frac{\operatorname{tr}(\Pi_c)}{d_s} \operatorname{tr}\left[\frac{\Pi_c}{\operatorname{tr}[\Pi_c]} \Pi_c\right] = \sum_{c=1}^{n} \frac{1}{d_s} \operatorname{tr}\left(\Pi_c^2\right) . \tag{4.20}$$

Each Π_c can be written as its spectral decomposition

$$\Pi_c = \sum_{k=1}^{r_c} \mu_{c,k} |k_c\rangle \langle k_c| \tag{4.21}$$

where $r_c = \operatorname{rank}(\Pi_c)$ and $\mu_{c,k}$ are the eigenvalues of Π_c, with $|k_c\rangle \langle k_c|$ to be the corresponding projector. Substituting the spectral decomposition into (4.20) gives

$$\sum_{c=1}^{n} \frac{1}{d_s} \operatorname{tr}\left(\Pi_c^2\right) = \frac{1}{d_s} \sum_{c=1}^{n} \sum_{k=1}^{r_c} \mu_{c,k}^2 \tag{4.22}$$

But using the Cauchy-Schwarz inequality, we have

$$\left[\sum_{c=1}^{n}\sum_{k=1}^{r_c}\mu_{k,c}^2\right]\left[\sum_{c=1}^{n}\sum_{k=1}^{r_c}1^2\right] \geq \left[\sum_{c=1}^{n}\sum_{k=1}^{r_c}\mu_{c,k}\right]^2$$

that is,

$$\sum_{c=1}^{n}\sum_{k=1}^{r_c}\mu_{c,k}^2 \geq \frac{d_s^2}{\sum_{c=1}^{n} r_c}, \tag{4.23}$$

because $\sum_{c=1}^{n}\sum_{k=1}^{r_c}\mu_{c,k} = \sum_{c=1}^{n}\mathrm{tr}\left(\Pi_k\right) = d_s$. By substituting this inequality into (4.22), we find finally the following lower bound on the guessing probability:

$$G(\mathbb{1}/d_s, \{\Pi_c\}) \geq \frac{d_s}{\sum_{c=1}^{n} r_c} \tag{4.24}$$

that is, the upper bound on the min-entropy

$$H_{\min}(\mathbb{1}/d_s, \{\Pi_c\}) \leq \log(\sum_c r_c) - \log d_s\,. \tag{4.25}$$

In the tensor product implementation of the POVM, which uses an ancilla of dimension d_a, we have $\sum_c r_c = d_s\, d_a$; whence the maximum min-entropy is $\log d_a$, which comes solely from the ancilla. In the direct sum implementation, $\sum_c r_c = d_s + d_h$ is the minimum total dimension (system + hidden subspace) required to implement the POVM [100]: since $\log(d_s + d_h) \leq \log d_s + \log d_h$, the min-entropy is upper bounded by $\log d_h$. In both cases, we have proved our claim: all the randomness that can be obtained in a POVM on the maximally mixed state can be ascribed to the additional degrees of freedom used to implement the POVM. Finally notice that, as far as our proof goes, this conclusion applies only when the system is in the maximally mixed state: it remains an open problem whether, in other cases, POVMs may extract more randomness from the system than projective measurements.

4.1.4(c) Randomness from pointer measurements

The previous observation on POVMs extends to another case in which additional degrees of freedom are used: that of measurement by coupling the relevant degree of freedom to a pointer. The best known textbook example is the Stern-Gerlach experiment. More common nowadays is the *measurement of the polarization of a photon*: the photon is sent on a polarizing beam-splitter (PBS), the two output ports of which are correlated with orthogonal polarizations. It is by detecting in which beam the photon is (pointer) that polarization is inferred. Now, if one sets up an experiment to extract randomness from the polarization qubit [101,

102], the same setup provides another degree of freedom, whose state must be very close to pure if the measurement has to make sense at all: indeed, the beam must come from a well-defined direction for the PBS to work as expected. Thence, in principle one can extract more randomness by ignoring polarization and sending the photon on a normal beam-splitter [103, 104]. We stress that this argument bears on the amount of randomness and on simplicity "on paper": polarization may be preferable to deal with other practical concerns [101].

For other qubits, things may be more subtle. Consider for instance the probing of an atomic qubit with a laser beam: a laser beam alone can be used to generate randomness [105], but with a different detection scheme than the one used in probing atomic excitations; so it may not be immediate to suggest that one should ignore the atom and extract randomness directly from the laser. For yet other pointer measurements, it may not even be feasible to measure the pointer in a complementary basis (certainly it would be challenging for the Stern-Gerlach setup).

It is not our aim to propose concrete schemes to extract randomness at the tomography level of characterization. Rather, the bottom-line message could be put this way: whenever a quantum degree of freedom is measured by coupling it to a pointer, the pointer is usually in a well-defined quantum state. So, *if the goal is to extract randomness, it is worth considering the possibility of getting it directly from the pointer*.

4.1.5 Conclusions

In this work, we have quantified the randomness that can be extracted from a given quantum device with the device-independent, one-sided device-independent, and tomographic level of characterization. Specific tools were introduced to perform this quantification in the one-sided device-independent and tomographic cases. For the latter, we have also shown that not all conceivable procedures to extract randomness are actually relevant: in particular one must be careful whenever ancillas are involved since the randomness may just come from them rather than from the system under study. We have focused on the minimal class of adversarial power, relevant for the study of implementations performed by trusted experimentalists; a similar study could be conducted for randomness extraction against more powerful adversaries.

In the next Section, we study the consequences of allowing *measurement dependence*, which we do not assume for the results of this current Section.

4.2 The role of randomness in Bell tests

The violation of Bell inequalities have been used to certify important quantum information properties in a black-box scenario under minimal assumptions. This idea of "device-independent" certification started in the context of quantum key distribution, where the violation of Bell inequalities bounds the information leaked to the eavesdropper [27, 106, 107]; and it has been extended to various other tasks, notably state certification [106, 108, 109], measurement certification [110], and private randomness expansion [29, 111, 112]. Ultimately, this stems from the fact that the violation of Bell inequalities certifies the presence of a quantifiable amount of intrinsic randomness: indeed, a contrario, if the outcomes were predictable, one could have predicted them in advance and the measurement could consist of reading from a pre-existing list. This is exactly what the violation of Bell inequality certifies as impossible (provided locality is assumed).

Unfortunately, two assumptions are left in device-independent certification. The first is no-signaling: the choice of the measurement setting of one party should not be known to the measurement boxes of the other parties before they produce their outcome. This can be guaranteed ultimately by ensuring space-like separation, although one may also trust a weaker demonstration of separation, as for instance in [111]. The second assumption is measurement independence: the information λ contained in the boxes in each run should be uncorrelated from the choice of the settings in that run. So far, no way of checking measurement independence is known in a black-box scenario: the best one can do is to buy the source of λ and the devices that choose the settings from different providers, who are believed not to be conspiring together. Alternatively, one can partly give up the black-box scenario, characterize the devices and be confident that the relevant degrees of freedom are uncorrelated.

It is clear that no-signaling and measurement independence cannot be arbitrarily relaxed: if any amount of signaling is allowed, or if arbitrary correlation is admitted between source and settings, the violation of a Bell inequality can be obtained with purely classical resources λ, so there is no hope to conclude that λ contains intrinsic randomness. However, with the aim of reducing the assumptions of device-independent certification to their bare minimum, one can partially relax no-signaling and measurement independence, and ask how much information must be signaled and how much measurement dependence must be allowed for a Bell test to become irrelevant [84]. In this paper, we focus on the latter question, the study of partial measurement dependence (sometimes called reduced measurement independence or reduced "free will"), which has been the object of a few recent

studies [32, 33, 113, 114]. In particular, we consider the random source that is required to choose the input settings for a Bell inequality and place bounds on the min-entropy necessary to show any difference between local and no-signaling output distributions. Note that if the violation of a Bell inequality is used in a device independent protocol to certify the amplification or expansion of input randomness, this source would serve as the seed randomness in the protocol.

4.2.1 Measurement dependence and its basic consequences

4.2.1(a) Measurement independence

For the sake of this introduction, we consider a bipartite Bell scenario. Operationally, a Bell experiment consists of N apparently identical runs[3], in each of which box A receives input x and outputs a value a, box B receives input y and outputs a value b. A measurement-setting source (henceforth source) for the Bell test supplies the experimentalist with inputs x and y; its behavior is modeled by a probability distribution $p(xy|\lambda)$. One can then estimate the statistics $p(ab|xy)$. We denote by λ the information present in the boxes in a given run.

Measurement independence, the assumption that we want to relax, is captured by the condition

$$p(\lambda|xy) = p(\lambda)\ \forall x, y\,. \tag{4.26}$$

Under this assumption, the observed statistics are modeled by

$$p_{\mathrm{MI}}(ab|xy) = \int p(ab|xy\lambda)p(\lambda)\,\mathrm{d}\lambda\ . \tag{4.27}$$

The specific goal of a Bell test is to assess whether there is intrinsic randomness in the boxes, that is, in the usual terminology, to guarantee that λ is not a local variable. Mathematically, local variables are defined by $p(ab|xy\lambda) = p(a|x\lambda)p(b|y\lambda)$. It is useful to stress that, as written, (4.27) contains an additional assumption, namely that λ itself is chosen independently in each run according to the distribution $p(\lambda)$. Under measurement independence, it can be proved that this is ultimately not a restriction for Bell tests, although one has to be careful in interpreting statistics from finite samples [90, 91, 115].

Measurement independence cannot be denied in a systematic way without undermining the scientific method itself (if a clinical trial is to make sense, whether each patient receives the drug or the placebo cannot depend on the any details of

[3]We focus on the operational description of current experiments and do not consider the more general, but as yet abstract, case of parallel repetition, in which all the inputs are given at the same time.

the patients' conditions). However, it is certainly possible to question measurement independence in a given setup: the devices that determine the inputs x, y may be correlated to the process that determines λ. The origin of such correlation may be trivial, like the fluctuations in power of the city network to which all the devices are connected; it may be due to lack of attention of the experimentalists, who introduced unwanted connections; or it may be strongly conspiratorial, in an adversarial scenario in which the devices come from an untrusted provider. In all cases, (4.26) does not hold, nor does the proof that one can restrict the study to independently-chosen λ.

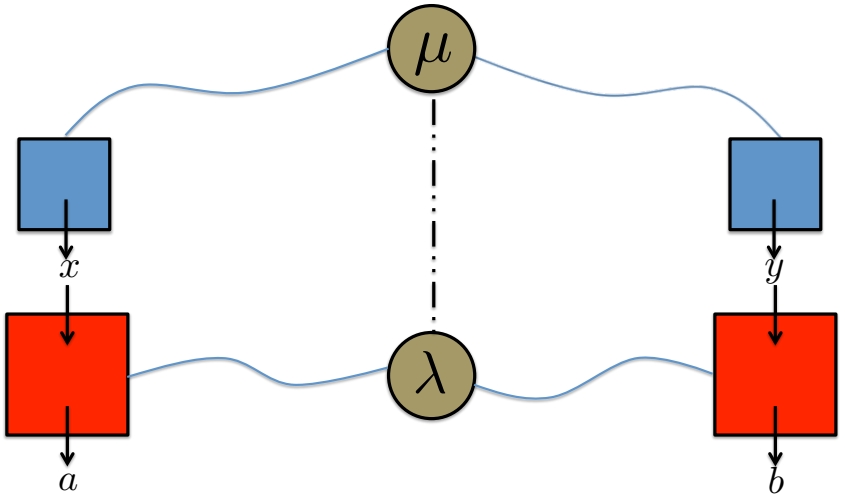

Figure 4.4: There are many different processes by which the information λ in devices A and B might become correlated with the inputs to the devices x and y, as discussed in the text. In this illustration the processes are represented by some external pre-existing variable μ that serves to introduce the correlation. The blue boxes represent the physical random number generators used to pick the inputs to the Bell test.

By relaxing condition (4.26), one allows correlations between the boxes' content λ and the choice of the settings x, y. Bayes theorem implies that

$$p(\lambda|xy) \neq p(\lambda) \Longleftrightarrow p(xy|\lambda) \neq p(xy)\,. \tag{4.28}$$

The first relation could be read as "the output of the source is restricted for a given choice of settings", the second as "the choice of settings is restricted for a given output of the source". Neither needs to refer to a real causal relation: all is compatible with both λ and x, y being influenced by a common cause (Figure 4.4). That being clarified, our discourse will be mostly phrased in the second way. We shall then look at measurement dependence as *reducing the probability of certain pairs of settings*. In the case where the dependence is sufficient to *exclude* enough

pairs of settings, unwanted features of local variable models may be hidden. This is the same intuition behind the power of the detection loophole; in fact, measurement dependence is even stronger, because it may allow to exclude a single *pair* of settings, whereas the detection loophole is local and excludes all pairs of settings such that one given setting of Bob, for instance, is associated to unwanted features. This opens a wealth of possibilities that we review rapidly next.

4.2.1(b) Effects of measurement dependence

The obvious effect of measurement independence is the possibility of *faking a violation of Bell inequalities*. A Bell inequality is built on a linear combination of $p(ab|xy)$, whose maximal value (called algebraic limit) cannot be reached by local variables. If, in each run, one can exclude some suitable pairs of settings in correlation with the content of the boxes λ, then it becomes possible to reach the algebraic limit while having only local variables in the boxes.

Let us illustrate this point with the most famous Bell inequality, the CHSH inequality

$$|\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle| \leq 2 \tag{4.29}$$

with $A_x, B_y \in \{-1, +1\}$. In order to achieve the algebraic limit of 4, one should have $A_0 = B_0$, $A_1 = B_0$, $A_0 = B_1$ and $A_1 = -B_1$. Local deterministic points exist that satisfy three out of these four conditions. If one wants to achieve the algebraic limit with local variable and measurement dependence, a sufficient strategy is the following: in each run, λ is chosen among the aforementioned local deterministic points, and the pair of settings corresponding to the unwanted condition is never chosen [84, 113].

The fact that a sufficient amount of measurement dependence can lead to the algebraic limit has an intriguing consequence for some inequalities. Indeed, in generic inequalities, the algebraic limit may lie even above what can be reached with no-signaling correlations. For instance, the tilted CHSH inequality

$$|\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle + \alpha\langle A_0\rangle| \leq 2 + \alpha \ , \tag{4.30}$$

has an algebraic limit of $4 + \alpha$, but no-signaling correlations can reach only up to 4 if $\alpha \leq 2$ [92]. If measurement dependence is allowed, to the point that one pair of settings can be excluded, then one can achieve the algebraic limit with a convex

mixture of

$$\begin{array}{llll} \lambda = (+1,-1,-1,+1) & \text{together with} & (x,y) \neq 00 \\ \lambda = (+1,+1,+1,-1) & \text{together with} & (x,y) \neq 01 \\ \lambda = (+1,-1,+1,+1) & \text{together with} & (x,y) \neq 10 \\ \lambda = (+1,+1,+1,+1) & \text{together with} & (x,y) \neq 11 \end{array} \tag{4.31}$$

where we denoted a local deterministic point as $\lambda = (A_0, A_1, B_0, B_1)$. If a Bell test is run with this underlying strategy, the observed correlations will lie outside the no-signaling polytope, i.e. are formally signaling. Obviously, this does not mean that measurement dependence makes it possible to use entanglement to actually send a message: in order for (say) Alice to send a message to Bob, she must be able to choose her setting at will, which is precisely what measurement dependence denies. At any rate, one must be careful when working with measurement dependence: the worst case are correlations that reach the algebraic limit, not the no-signaling one (to our knowledge, all the studies of measurement dependence so far dealt with inequalities for which the two limits happen to coincide [32, 33, 84, 113, 114]).

The take-away message of this paragraph is that one does not have to reach the extreme case of total measurement dependence (i.e. λ determining x, y uniquely): *already with some partial amount of measurement dependence, it becomes impossible to draw any conclusion from the violation of a Bell inequality.* This has important consequences when the source is characterized only by its conditional min-entropy. Indeed, one of our main result will consist in deriving general bounds for this amount (Section 4.2.3). In order to do that, we need first to recall the definition of min-entropy and its relation to the Santha-Vazirani condition in light of measurement dependence.

4.2.2 Min-entropy and measurement dependence

As mentioned, the source of the Bell test behaves according to $p(xy|\lambda)$. Measurement independence implies that $p(xy|\lambda)$ has as much entropy or randomness as $p(xy)$. In contrast, partial measurement dependence means that there is some randomness in the source, but it is less than the entropy of the distribution $p(xy)$. The min-entropy and min-entropy deficit $H_{\min}(\mathbf{Z}) - H_{\min}(\mathbf{Z}|\mathbf{\Lambda})$, where $\mathbf{Z}$ and $\mathbf{\Lambda}$ are strings over some alphabet, are measures of randomness of a source, and they partly capture the amount of measurement dependence in special cases. But note that they are not intrinsic measures of measurement dependence (for instance, min-entropy deficit equals zero does not imply measurement independence). If the min-entropy is not high enough, it leaves open the possibility of excluding

certain settings, which allows faking of Bell violations as we discussed before. This behavior is forbidden in Santha-Vazirani sources as explained next.

4.2.2(a) Min-entropy vs Santha-Vazirani condition

We illustrate our point with an example. The *chained inequality* is a bipartite Bell inequality with m settings for each party and binary outcomes a, b for both measurements on A and B, which reads

$$I_m = p(a = b|x = 1, y = m) + \sum_{\substack{x,y \text{ s.t.} \\ x \in \{y, y+1\}}} p(a \neq b|x, y) \leq 2m - 1 \,. \tag{4.32}$$

It has been used to put stringent bounds on quantum theory thanks to the property that, in the limit $m \to \infty$, its algebraic limit $I_m = 2m$ can be reached with measurements on quantum states [116, 117].

Out of the m^2 possible pairs of settings, $2m$ are effectively used in the inequality. Furthermore, there exist local deterministic points that can satisfy $2m - 1$ of these conditions. Therefore, in order to verify any conclusion based on the chained inequality, it is enough to have an amount of measurement dependence that allows the exclusion of only one pair of settings out of m^2. In the limit of large m, under whichever measure, such a source is very close to a fully random source: for instance, its min-entropy per run (defined below) is $\log(m^2 - 1)$, which differs from the fully random value $\log m^2$ by $O(m^{-2})$. This example shows that a source, which would presumably be considered as good as it gets in an abstract assessment, is already catastrophic for the Bell inequality under study. Notice that this remark is not in contradiction with the results of [32], which can be seen as proving that the chained inequality is pretty robust to measurement dependence: indeed, in that work, the additional Santha-Vazirani assumption was made on the source, which implies that all the pairs of settings are possible in each run. Our argument, based on excluding one setting in each run, does not apply.

It is now time to present the definitions we have just sketched in their suitable formal setting. We shall consistently use the word *source* to stress that the source of randomness we are interested in is the randomness of the inputs given the knowledge of the physical process λ or vice versa, not the randomness possibly present in λ (which would be the intrinsic randomness of quantum origin in the ideal case).

4.2.2(b) Formal definitions

Here we review rapidly the definitions of well-known types of sources of randomness for the purpose of this paper, referring to [118] for a comprehensive study.

Consider a random variable Z in an alphabet $\mathcal{Z}$ of size d; and let $\mathbf{Z} = Z_1...Z_N$ be an N-dit string. In our case, Z will represent the settings chosen for the Bell test, i.e. $Z = (x, y)$ in a bipartite scenario. Randomness being synonymous with unpredictability, a source of randomness will be characterized by specifying what one wants to predict and how predictable it is, given some prior information $\mathbf{\Lambda}$ (supposed to be classical throughout this paper). One would then say that the source contains randomness if (see also the preliminary chapter)

$$P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda}) := \sum_{\lambda} P(\mathbf{\Lambda} = \lambda) P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda) < 1\,, \tag{4.33}$$

where $P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda) := \max_{\mathbf{z}} p(\mathbf{Z} = \mathbf{z}|\mathbf{\Lambda} = \lambda)$. The amount of randomness is quantified by the min-entropy

$$H_{\min}(\mathbf{Z}|\mathbf{\Lambda}) := -\log P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda})\,. \tag{4.34}$$

Clearly, $H_{\min}(\mathbf{Z}|\mathbf{\Lambda}) > 0$ implies the presence of some randomness. To someone who does not have access to $\mathbf{\Lambda}$, the source will appear to have min-entropy $H_{\min}(\mathbf{Z}) = -\log P_{\text{guess}}(\mathbf{Z})$ which can only be higher by the data processing inequality. Though obvious, it may be worth stressing that $\sum_\lambda P(\mathbf{\Lambda} = \lambda) P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda)$ is *not* the same as $P_{\text{guess}}(\mathbf{Z})$, since P_{guess} is not a given probability distribution but a notation for a procedure that picks up the maximum of a probability distribution. As an extreme example, if $\mathbf{Z}$ looks uniform but the knowledge of $\mathbf{\Lambda}$ determines $\mathbf{z}$ uniquely, one has $P_{\text{guess}}(\mathbf{Z}) = 1/d^N$ and $\sum_\lambda P(\mathbf{\Lambda} = \lambda) P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda) = 1$.

The loosest characterization of the source, i.e. the one that requires fewer assumptions, simply puts a bound on the min-entropy:

Definition 5. Min-entropy source. *A random variable* $\mathbf{Z}$ *is a* k *min-entropy source of randomness with respect to another random variable* $\mathbf{\Lambda}$ *if* $H_{\min}(\mathbf{Z}|\mathbf{\Lambda}) \geq k$.

As soon as $k > 0$, the knowledge of $\mathbf{\Lambda}$ does not determine $\mathbf{z}$ uniquely. One can add some structure to a min-entropy source. For instance, a k-min-entropy source is called *uniform* if $H_{\min}(\mathbf{Z}|\mathbf{\Lambda} = \lambda) := -\log P_{\text{guess}}(\mathbf{Z}|\mathbf{\Lambda} = \lambda) \geq k$ for all values of λ. A *block min-entropy source* is one for which not only the min-entropy of the whole string, but the min-entropy of blocks is also lower bounded.

As soon as $k \leq \log(d^N - 1)$, the definition of k-min-entropy source is compatible with $P_{guess}(\mathbf{Z} = \mathbf{z}|\mathbf{\Lambda} = \lambda) = 0$ for one string $\mathbf{z}$. As hinted in **Min-entropy vs**

Santha-Vazirani condition, the possibility that some settings are not chosen is critical for sources of Bell tests. Because of this, one may want to add to the properties of the source the assumption that *all* the d^N strings have non-zero probability. This is equivalent to the following type of source:

Definition 6. *Santha-Vazirani sources. A random variable* $\mathbf{Z}$ *is a* (p_{min}, p_{max}) *Santha-Vazirani source with respect to* $\mathbf{\Lambda}$ *(where* $0 \leq p_{min} \leq 1/d$ *and* $1/d \leq p_{max} \leq 1$*) if*

$$p_{\min} \leq p(z_i|\lambda, z_1, ..., z_{i-1}) \leq p_{max} \quad \forall\ i\ . \tag{4.35}$$

If Z_i is a bit, $p_{\min} = 1 - p_{\max}$ is usually written δ [119]. Some of the most important results in measurement dependence in Bell tests have been obtained for Santha-Vazirani sources [32, 33, 114]. These results show that there is a real advantage in considering Bell-based randomness, because it overcomes no-go theorems for classical information.

Finally, let us focus on distributions that are independent and identically distributed (i.i.d.) such that

$$p(\mathbf{Z} = \mathbf{z}|\mathbf{\Lambda} = \lambda) = \prod_{j=1}^{N} p(Z_j = z_j|\lambda). \tag{4.36}$$

This can also be viewed as a block min-entropy source where each block consists of only one symbol, Z_i. In this case, the Santha-Vazirani definition implies:

$$p_{\min} \leq p(z|\lambda) \leq p_{max}\ . \tag{4.37}$$

We will use a different notation such that $p_{\max} = P_M$ and $p_{\min} = P_m$ to make clear that we are in the i.i.d. scenario. Then the definition of uniform min-entropy sources is equivalent to the figure of merit of measurement dependence used in [113], namely

$$P_M := \max_{z,\lambda} p(z|\lambda)\ , \qquad \text{[i.i.d.]} \tag{4.38}$$

since $H_{\min}(Z|\Lambda = \lambda) \geq k$ for all λ is equivalent to $P_M \leq 2^{-k}$. In the following, we will use these two figure of merits interchangeably for i.i.d. models.

Instead of bounding the largest probability, the smallest probability also gives information on measurement dependence, as first proposed in [120]:

$$P_m := \min_{z,\lambda} p(z|\lambda)\ . \qquad \text{[i.i.d.]} \tag{4.39}$$

If only P_M is explicitly bounded, then a bound on P_m can be inferred, however, it might be trivial, since it can be negative: $P_m \geq 1 - (d-1)P_M$. Bounding only the

min-entropy of the input source to the Bell test, or equivalently bounding only P_M, which is the guessing probability, allows much different worst-case behavior in Bell tests than when the Santha-Vazirani definition is adopted, as we shall now explore.

4.2.3 Lower bound for min-entropy sources

We will be dealing with a K-partite Bell scenario where the i^{th} party has $m_i > 1$ measurement settings (m_A, m_B for bipartite) and each setting has an arbitrary number of outcomes. The joint configuration of settings $\mathbf{z} = z_1...z_K$ with $z_i \in \{1, ..m_i\}$ ($\mathbf{z} = xy$ for bipartite) is a K-tuple in the set of all settings $\mathcal{S}$ of size $\prod_{i=1}^{K} m_i$. In this Section, moreover, we consider a Bell test in which the observed statistics of the settings follow a uniform distribution, that is

$$H_{\min}(\mathbf{Z_1}, ..., \mathbf{Z_K}) = N \log |\mathcal{S}| \,, \tag{4.40}$$

or equivalently

$$p_{obs}(z_1...z_K) := \sum_{\lambda} p(z_1...z_K|\lambda)p(\lambda) = \Big(\prod_{i=1}^{K} m_i\Big)^{-1}. \tag{4.41}$$

This is not an assumption like those on the nature of the source: p_{obs} is observed in a realization; but it is a frequent working assumption for theoretical works, which was made in all previous works on measurement dependence. In Section 4.2.4, we shall see that a non-uniform p_{obs} has interesting consequences in studies of measurement dependence.

We are presently able to discuss our main result: a lower bound on the min-entropy of the source, below which no conclusion can be drawn from any Bell test, unless further structure is assumed.

4.2.3(a) Reaching the no-signaling limit

The main insight is provided by the following Lemma, which we present in the bipartite scenario, and generalize to multipartite scenarios. Intuitively, the lemma can be paraphrased as follows: by changing the output behavior for some input (i.e. for some xy, changing the distribution $p(ab|xy)$) any no-signalling correlation can be made local.

Lemma 1. *For any pair of settings $(\bar{x}, \bar{y})$, and for all $p(ab|xy)$ being an arbitrary no-signaling correlation with $x \in \{1, ..., m_A\}$ and $y \in \{1, ..., m_B\}$, there exists a*

local distribution $p_L(ab|xy)$ *such that*

$$p_L(ab|xy) = p(ab|xy) \quad (4.42)$$
$$\text{for } (x,y) \in \mathcal{S}_{\bar{x},\bar{y}} \equiv \{(\bar{x},y'),(x',\bar{y}) : x' \in \{1,...,m_A\}, y' \in \{1,...,m_B\}\}.$$

Moreover, this result is tight: if another pair of settings is added to the subset of pairs, there exists a no-signaling point for which those probabilities are nonlocal.

Proof. The proof can be done by constructing explicitly one such local distribution. Let us fix $(\bar{x},\bar{y}) = (1,1)$ without loss of generality. From the no-signaling distribution p, we construct

$$\boldsymbol{p}(a_1,a_2,...,a_{m_A};b_1,b_2,...,b_{m_B}) = p(a_1)p(b_1|a_1)\prod_{j=2}^{m_A} p(a_j|b_1)\prod_{k=2}^{m_B} p(b_k|a_1) \quad (4.43)$$

with obvious notations for the marginals. This is a valid joint probability distribution over the outcomes of all the measurements. Now, on the one hand, the marginals $\boldsymbol{p}(a_j;b_k) \equiv p_L(a,b|j,k)$ define a local distribution, as first proved by Fine [121]. On the other hand, it is easy to show that $\boldsymbol{p}(a_1;b_k) = p(a,b|1,k)$: one should sum first over all possible values of $a_2,...,a_{m_A}$ to find $\boldsymbol{p}(a_1;b_1,b_2,...,b_{m_B}) = p(a_1)\prod_{k=1}^{m_B} p(b_k|a_1)$, after which the sum over the b's is obvious. Similarly one proves that $\boldsymbol{p}(a_j;b_1) = p(a,b|j,1)$. So indeed we have a local distribution that mimicks the initial no-signaling one on the desired subset of pairs of settings.

As for the tightness, suppose that we add a single pair of settings, say $(2,2)$, to $\mathcal{S}_{1,1}$: there exist no-signaling points for which CHSH is violated by the settings $(1,1)$, $(1,2)$, $(2,1)$ and $(2,2)$; so those statistics can't be mimicked by a local distribution. □

Lemma 2. *For any* K*-tuple of settings* $\bar{\boldsymbol{z}} = (\bar{z}_1,...,\bar{z}_K)$ *and for all* $p(\boldsymbol{o}|\boldsymbol{z})$ *being an arbitrary* K*-partite no-signaling correlation with* $\boldsymbol{z} = (z_1,...,z_K)$ *where* $z_i \in \{1,...,m_i\}$ *and* $\boldsymbol{o}$ *is a* K*-tuple of outcomes, there exists a local distribution* $p_L(\mathbf{o}|\mathbf{z})$ *such that*

$$p_L(\boldsymbol{o}|\boldsymbol{z}) = p(\boldsymbol{o}|\boldsymbol{z}) \quad (4.44)$$
$$\text{for } \boldsymbol{z} \in \mathcal{S}_{\bar{\boldsymbol{z}}} \equiv \left\{(\bar{z}_1,z_2',...,z_K'),...,(z_1',...,z_{K-1}',\bar{z}_K) : z_i' \in \{1,...,m_i\}\right\}.$$

Moreover, this result is tight: if another K*-tuple of settings is added to the subset* $\mathcal{S}_{\bar{\mathbf{z}}}$*, there exist a no-signaling point for which those probabilities are nonlocal.*

Proof. Again, let us fix $\bar{\boldsymbol{z}} = (1,...,1)$ without loss of generality. Let $o^i_{z_i}$ be the ith party's outcome given the z_i^{th} measurement setting. From the no-signaling

distribution p, we construct a valid probability distribution

$$\boldsymbol{p}(o_1^1...o_{m_1}^1;...;o_1^K...o_{m_K}^K) \tag{4.45}$$

$$= p(o_1^1)\left[\prod_{i=2}^{K} p(o_1^i|o_1^1...o_1^{i-1})\right]\left[\prod_{i=1}^{K}\prod_{j=2}^{m_i} p(o_j^i|o_1^1...o_1^{i-1}o_1^{i+1}...o_1^K)\right] \tag{4.46}$$

whose marginals $\boldsymbol{p}(o_{z_1}^1;...;o_{z_K}^K) \equiv p_L(o^1...o^K|z_1...z_k)$ define a local distribution by Fine's result [121]. To verify that we have a local distribution that mimics the initial no-signaling one on the desired subset of pairs of settings, consider this example: for the input string $(1, z_2, ...z_k)$ with the distribution $\boldsymbol{p}(o_1^1; o_{z_2}^2; ...; o_{z_K}^K) = p_L(o^1o^2...o^K|1, z_2, ...z_k)$ we sum first over all possible values of each outcome variable $o_2^1, ..., o_{m_1}^1$ to find

$$\boldsymbol{p}(o_1^1; o_1^2...o_{m_2}^2; ...; o_1^K...o_{m_K}^K) \tag{4.47}$$

$$= p(o_1^1)\prod_{i=2}^{K}\left[p(o_1^i|o_1^1...o_1^{i-1})\prod_{j=2}^{m_i} p(o_j^i|o_1^1...o_1^{i-1}o_1^{i+1}...o_1^K)\right] \tag{4.48}$$

after which continue to sum over all the $o_k^2, ..., o_k^K$ except $o_{z_2}^2, ..., o_{z_K}^K$ and one is left with a probability distribution $\mathbf{P}(o_1^1, o_{z_2}^2...o_{z_K}^K)$ on only K variables, one for each party. The other verifications are similar. Another way to think of it is to notice that each conditional probability factor on K variables (one variable conditioned on $K-1$ other variables) effectively sets a joint probability distribution on those same K variables. In the distribution (4.46) there are $\sum_{i=1}^{K} m_i - K + 1$ such factors and so this is exactly how many local points $p_L(\boldsymbol{o}|\boldsymbol{z})$ that can be matched for a given hidden variable value (see equation (4.50) in the main text). The argument for tightness still works if we consider only two parties among K. For any two parties we can choose a pair of inputs for each to return to a CHSH-type scenario, then the argument follows in the same way as in the proof of Lemma 1.

□

Now we can state the main theorem:

Theorem 2. *Consider a min-entropy source with an observed min-entropy* $H_{\min}(\mathbf{XY}) = N\log(m_Am_B)$ *for an* N*-run bipartite Bell test with* m_A *inputs on Alice,* m_B *inputs on Bob and arbitrary alphabets for the outcomes. If*

$$H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) \leq N\ \log(m_A + m_B - 1) \tag{4.49}$$

no conclusion can be drawn from the Bell test, since the no-signaling limit of the inequality can be reached with local distributions. The generalization of this result

to K-partite Bell tests reads

$$H_{\min}(\mathbf{Z_1}...\mathbf{Z_K}|\mathbf{\Lambda}) \leq N \log\Big(\sum_{k=1}^{K} m_k - K + 1\Big). \tag{4.50}$$

for $H_{\min}(\mathbf{Z_1}...\mathbf{Z_K}) = -\log(\prod_K m_K)$. Notice in particular that, without further assumptions, any source of randomness with $H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) \leq N \log 3$ is useless as a source for any Bell tests.

Proof. We will construct an explicit i.i.d. source which allows the faking of a Bell violation up to the no-signaling bound with appropriate local resources. From Lemma 1 we know that there exist subsets $\mathcal{S}_{\bar{x},\bar{y}}$ of $m_A + m_B - 1$ pairs of settings, for which no difference can be seen if a local distribution is substituted for a possibly nonlocal no-signaling point: in particular, this could be the no-signaling point that reaches the no-signaling limit for the inequality under study. If $H_{\min}(\mathbf{XY}|\mathbf{\Lambda})$ is sufficiently low, the source will allow only the pairs of settings that belong to one of the $\mathcal{S}_{\bar{x},\bar{y}}$ and distribute the corresponding local strategy $\lambda_{\bar{x},\bar{y}}$. The source

$$p(xy|\lambda_{\bar{x},\bar{y}}) = \begin{cases} \frac{1}{m_A+m_B-1}, & \text{if } x,y \in \mathcal{S}_{\bar{x},\bar{y}} \\ 0, & \text{otherwise} \end{cases} \tag{4.51}$$

has $P_M = \frac{1}{m_A+m_B-1}$ in each run, whence we have proved the bound (4.49) as long as we can find $p(\lambda_{\bar{x},\bar{y}})$ such that $\sum_{\bar{x},\bar{y}} p(xy|\lambda_{\bar{x},\bar{y}})p(\lambda_{\bar{x},\bar{y}}) = p_{obs}(xy)$ for all x, y. In the case where p_{obs} is uniform, this can always be found by simply choosing uniformly the pair $(\bar{x},\bar{y})$, i.e. $p(\lambda_{\bar{x},\bar{y}}) = \frac{1}{m_A m_B}$. This concludes the proof for the bipartite case. The proof of the multipartite case is identical using the material of Lemma 2. The final remark of Theorem 1 stems from the fact that each Bell test much involve at least two parties and each must have at least two settings.

□

Two remarks can be drawn from our result at this point. The first is on randomness amplification. Given a single min-entropy source producing a string $\mathbf{z}$, without an additional seed it is well known that one cannot amplify the source classically: for any function f, there exists a min-entropy source $\mathbf{Z}$ such that $f(\mathbf{Z})$ is constant. Here we extend this impossibility result to the case of randomness amplification via no-signalling resources. One obstacle stands in our way: how to we connect the output from the source with the inputs to the Bell test? As in the classical case, without an additional independent seed, all one can do is to use the min-entropy source to pick the settings of an N run Bell test. This is formalized by applying a function from the alphabet of $\mathbf{Z}$ to the set of all possible setting

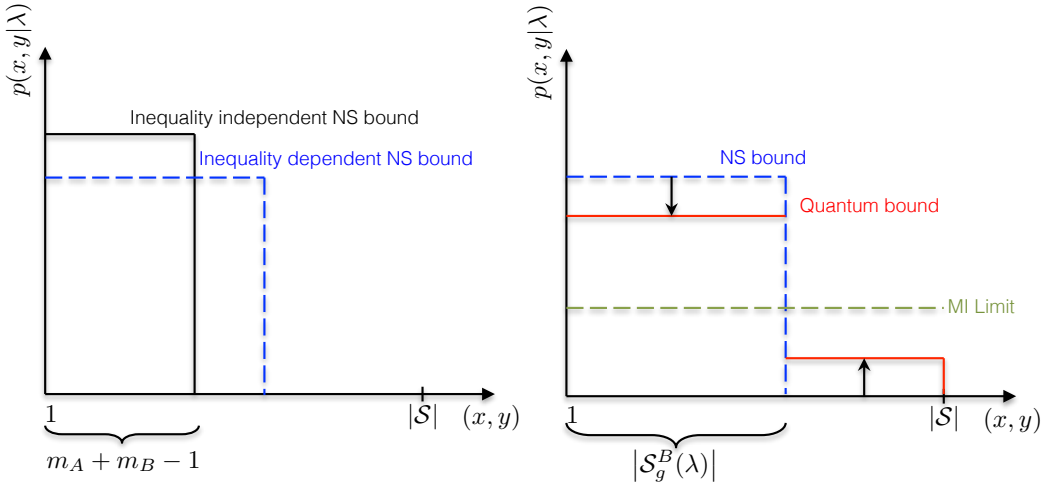

Figure 4.5: Sources that reach the critical min-entropy bound for uniform observed distribution of settings. For the inequality independent bound (4.50), the source is uniform on $m_A + m_B - 1$ settings and is zero elsewhere. For a given inequality, the no-signaling limit may be reached with a source that is uniform on a larger number of settings $|\mathcal{S}_g^B|$, and still zero on the others; in order to reach only the quantum limit, one can allow the settings in $\mathcal{S}_h^B$ to be used sometimes. Of course, for each λ, the settings that are chosen may vary.

combinations for the Bell test. Since this process cannot increase the min-entropy of the source, our main result extends to this case: one cannot amplify an arbitrary min-entropy source using no-signalling resources (in the device-independent level of characterization). Either the initial quality needs to be high enough or the level of characterization must increase (e.g. tomography level).

The second remark is about the critical min-entropy. Because of the tightness of Lemma 1, the bounds (4.49) and (4.50) are the best inequality-independent bounds that one can obtain with i.i.d. sources. Moreover, since there exist inequalities for which the quantum and the no-signaling limits coincide, the bound to reach the quantum limit cannot be better. If the inequality is given, however, much less measurement dependence may be sufficient to reach the no-signaling limit, and even less to reach the quantum limit if it is lower. We elaborate further on this point in the following paragraph.

4.2.3(b) Inequality-dependent bounds

Let B define a Bell inequality, whose local, quantum and no-signaling limits are given by $B_{\mathrm{L}} \leq B_{\mathrm{Q}} \leq B_{\mathrm{NS}}$, and $\mathcal{S}^B$ be the set of settings that are used by the Bell inequality [4]. Again, for each λ there is a local strategy for assigning outputs such

[4]The chained Bell inequality is an example of an inequality whose value depends only some of the possible inputs. Only terms such that $x = y$ or $x = y + 1$ and the term for $x = 1,\ y = m$ appear.

that in order to achieve the no-signaling limit, some settings will be incompatible with this strategy and must be hidden by measurement dependence. Let this set of inputs be $\mathcal{S}_h^B(\lambda)$. Then, an arbitrary no-signaling point is required to be compatible with a local point only on the subset $\mathcal{S}_g^B(\lambda) := \mathcal{S}^B \setminus \mathcal{S}_h^B(\lambda)$. Suppose an inspection of the inequality B shows that *at most* $|\mathcal{S}_h^B|$ of these $|\mathcal{S}^B|$ settings must be hidden for any choice of λ. Once the probabilities of the settings $\mathcal{S}_h^B(\lambda)$ are set to zero, the min-entropy is maximized by the uniform distribution over the remaining $|\mathcal{S}_g^B|$ settings (Figure 4.5). However, one must be very careful to show the existence of $p(\lambda)$ which satisfies (4.41). Whenever such a distribution exists, if $P_M \geq 1/|\mathcal{S}_g^B|$, the non-local game can be won with probability one with local strategies. As implied by the results of the previous section, if the observed input distribution $p_{obs}(\mathbf{z})$ is uniform then a strategy in the form of equation (4.51) or its generalization in Lemma 2 with a uniform probability over λ will always satisfy (4.41). However, it is possible to do better in some cases where $|\mathcal{S}^B| < |\mathcal{S}|$. In such cases $|\mathcal{S}_h^B|$ (the most settings that must be hidden for any λ) can be small. Here, for a uniform p_{obs} equation (4.41) will also be satisfied provided possibly more settings than required are hidden for each λ such that $|\mathcal{S}_h^B(\lambda)| = |\mathcal{S}_h^B|$ for all λ and if $\mathcal{L}(\mathbf{z})$ is the set of λs for which $\mathbf{z} \in \mathcal{S}_h^B(\lambda)$, $|\mathcal{L}(\mathbf{z})| = |\mathcal{S}_h^B|$ must be constant for all $\mathbf{z}$. This is a symmetry condition that can be met by many Bell inequalities. As before, the existence of this example proves that a min-entropy source with

$$k \leq N \log\left(|\mathcal{S}_g^B|\right) \tag{4.52}$$

can reach the no-signaling limit of B with local strategies for uniform input distributions. In the following section we will show how to obtain bounds for arbitrary p_{obs} and that approach will also give tight bounds and optimal strategies when the inequality is one in which the size of the "hidden sets" varies with λ.

Further, if $B_{\mathrm{Q}} < B_{\mathrm{NS}}$, in order to simulate physics one may be content with reaching the quantum limit. A possible i.i.d. source (not proved to be optimal) is the following (see Figure 4.5). With probability $1 - q$, the settings are chosen uniformly among all $\mathcal{M}$ possible K-tuples: this is measurement independence, so $B \leq B_L$ on these cases, and the physical process λ can be chosen as one of those that saturate $B = B_{\mathrm{L}}$. In the other instances, the settings are chosen uniformly in $\mathcal{S}_g^B$ and the physical process λ is chosen in each case in order to achieve $B = B_{\mathrm{NS}}$. In other words, this source is a convex combination of the measurement independent uniform source and the source described in the previous paragraph. Note that this new source will automatically satisfy the constraint (4.41). For such a source, therefore, P_M is the probability of each setting in $\mathcal{S}_g^B$, which reads

$P_M = \frac{1}{|\mathcal{S}|}\left[1 + q\left(|\mathcal{S}|/|\mathcal{S}_g^B| - 1\right)\right]$. With this measurement-dependent strategy, one can reach $B = qB_{\mathrm{NS}} + (1-q)B_L$, so $B \geq B_{\mathrm{Q}}$ for $q \geq (B_{\mathrm{Q}} - B_{\mathrm{L}})/(B_{\mathrm{NS}} - B_{\mathrm{L}})$. In summary, the quantum limit can be achieved with an i.i.d. source with

$$P_M^{B,Q} \geq \frac{1}{|\mathcal{S}|}\left[1 + \frac{B_{\mathrm{Q}} - B_{\mathrm{L}}}{B_{\mathrm{NS}} - B_{\mathrm{L}}}\left(|\mathcal{S}|/|\mathcal{S}_g^B| - 1\right)\right], \tag{4.53}$$

that is, a min-entropy source with $k \leq -N \log P_M^{B,Q}$ can reach the quantum limit of B with local strategies, for a uniform input distribution.

Let us illustrate the methodology with the analysis of some inequalities:

- *CHSH*: here, it is always necessary and sufficient to hide one pair of settings. Therefore $|\mathcal{S}_g^B| = 3$ and the inequality-dependent bound (4.52) is the same as the inequality-independent one (4.49) to reach the no-signaling limit, as already proved in [113]. Recall that this does not prove the bounds to be tight, because they are based on explicit i.i.d. sources: non i.i.d. sources may lead to tighter bounds, though we do not know any example. As for reaching the quantum limit, we have $P_M^{B,\mathrm{Q}} = \frac{1}{4}[1 + (\sqrt{2} - 1)/3] \approx 0.2845$.

- *Chained inequality*: here again, as we have seen in 4.2.2(a), it is always necessary and sufficient to hide only one pair of settings out of $\mathcal{M} = m^2$, so $|\mathcal{S}_g^B| = m^2 - 1$ and $|S_h^B(\lambda)| = 1$ for all λ. As a consequence, in terms of min-entropy, the inequality-dependent bound (4.52) is $N \log(m^2 - 1)$, which is approximately twice the value $N \log(2m - 1)$ obtained from (4.49). For large m, the quantum and no-signaling limits basically coincide.

- *CGLMP inequalities*: like the CHSH inequality, the CGLMP inequalities are two party inequalities where each party has two inputs. However, this family of inequalities has d possible outputs for each party. In the quantum case, the CGLMP inequalities can provide more robustness against measurement dependence than the CHSH inequality, in the sense that the min-entropy of the inputs given the source must be lower if the quantum bound is to be achieved. The reason is that it has been shown that as $d \to \infty$, the quantum limit increases and approaches the no-signaling limit [122, 123]. As can be seen, inspecting equation (4.53), the value of $(B_{\mathrm{Q}} - B_{\mathrm{L}})/(B_{\mathrm{NS}} - B_{\mathrm{L}})$ will increase with d, and the value of P_M necessary to reach the quantum limit with local resources increases, until it reaches the no-signaling value $P_M^{B,\mathrm{NS}}$ in the limit.

- *Mermin inequalities*: Mermin inequalities [124, 125] are multipartite inequalites such that for odd numbers of parties, the quantum and no-signaling

bounds coincide. For this reason, the 5-party Mermin inequality was used in [33] to amplify randomness. When the number of parties is an odd number at least 3 only a subset of all possible inputs appear in the corresponding Mermin inequality and the inequality-independent bound is not tight. In general, for odd K parties $|\mathcal{S}_g^B| = 2^{K-2} + 2^{(K-3)/2}$ and $|\mathcal{S}^B| = 2^{K-1}$ [126]. Specifically for the 5-party case, $|\mathcal{S}_g^B| = 10$ and $|\mathcal{S}^B| = 16$.

4.2.4 Counteracting measurement dependence

Theorem 1 shows that assuming a full min-entropy source on the measurement settings, for any meaningful conclusion to be drawn from a Bell test, it must be that $H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) > N \log(m_A + m_B - 1)$. However, recalling that the role of the observed data is actually a *constraint* imposed on the underlying model (similar to (4.41)), we can hope to use it to our advantage. This motivates the question: for a given value of $H_{\min}(\mathbf{XY}|\mathbf{\Lambda}) = Nk > N \log(m_A + m_B - 1)$ that is being assumed, what is the optimal distribution on the inputs such that the maximum possible Bell value obtainable with this degree of measurement dependence and only local resources is as low as possible. Because the situation for non i.i.d. models is intractable, we are restricting ourselves to the i.i.d. model for the remaining of this chapter. Here instead of the min-entropy, the guessing probability P_M is used exclusively as the figure of merit of measurement dependence. First, we consider the CHSH inequality as an explicit example.

4.2.4(a) The CHSH Inequality

Intuitively, we expect that the optimal solution is to set for each input round $H_{\min}(XY) = H_{\min}(XY|\Lambda) = k$ and $p_{obs}(xy) = 2^{-k}$ for three pairs (x, y) and $p_{obs}(x'y') = 1 - 3 \times 2^{-k}$ for the final pair because in this case Λ cannot contain any further information on XY than is available simply from observing the distribution $p_{obs}(xy)$. We will highlight an example of this type of distribution later in this section. This is not a uniform distribution, so we can already see that non-uniform input distributions can be beneficial. In this section, we will consider fixed input distributions $p_{obs}(xy)$ and find the maximum value that the CHSH inequality can take given a bound on P_M. Note that the method in this section extends to any multipartite Bell inequality.

We want to find the violation $B_{\text{CHSH}}^{\max}$, under local resources and measurement dependence, as a function of P_M and $p_{obs}(xy)$. To this end, observe that the local distributions form a convex polytope and so is the set of sources with a fixed value of P_M (the source polytope). Using the decomposition into extremal points of a

convex polytope, we have

$$p(ab|xy\lambda) = \sum_i \alpha_i(\lambda) e_i(ab|xy)\,, \tag{4.54}$$

$$p(xy|\lambda) = \sum_j \beta_j(\lambda) f_j(xy)\,, \tag{4.55}$$

where $e_i(ab|xy)$ are the extremal points of the local polytope and $f_j(xy)$ are the extremal points of the source polytope. Now after multiplying by both sides by $p_{obs}(xy)$, the i.i.d. model with measurement dependence becomes

$$p(abxy) = \int_\Lambda d\lambda \sum_{ij} \alpha_i(\lambda)\beta_j(\lambda) e_i(ab|xy) f_j(xy) p(\lambda) = \sum_{ij} \gamma_{ij} g_{ij}(abxy)\,, \tag{4.56}$$

where

$$\gamma_{ij} = \int_\Lambda d\lambda \alpha_i(\lambda)\beta_j(\lambda) p(\lambda)\,, \tag{4.57}$$

$$g_{ij}(abxy) = e_i(ab|xy) f_j(xy)\,. \tag{4.58}$$

In this notation, the problem becomes a linear program, i.e. finding

$$B^{\max}_{\text{CHSH}}(P_M, p_{obs}(xy)) = \max_{p(abxy)} \sum_{abxy} (-1)^{a+b+xy} \frac{p(abxy)}{p_{obs}(xy)} \tag{4.59}$$

subjected to the constraints

$$p(abxy) = \sum_{ij} \gamma_{ij} g_{ij}(abxy),\ \sum_{ab} p(abxy) = p_{obs}(xy) \tag{4.60}$$

for known values of $p_{obs}(xy)$ and P_M. The result is presented in Figure 4.6.

Using the numerical results, it is easy to see that the optimal strategy for maximizing the Bell value B_{CHSH} whether or not the observed distribution is uniform is to choose

$$p(xy|\lambda_{\bar{x},\bar{y}}) = \begin{cases} P_M & \text{if } x, y \in \mathcal{S}_{\bar{x},\bar{y}} \\ 1 - 3P_M & \text{otherwise} \end{cases}\,, \tag{4.61}$$

for each $\lambda_{\bar{x},\bar{y}}$, where $\mathcal{S}_{\bar{x},\bar{y}}$ is defined in equation (4.42). Choosing this strategy, it is straightforward to find an analytic expression for $B^{\max}_{\text{CHSH}}$:

$$B^{\max}_{\text{CHSH}}(P_M) = 4 - \frac{1}{2}\left[(1 - 3P_M)q + \frac{1 - 3P_M}{4P_M - 1}(q - 16)\right]$$

where for convenience we define $q := \sum_{x,y} \frac{1}{p_{obs}(x,y)}$. This expression is only valid

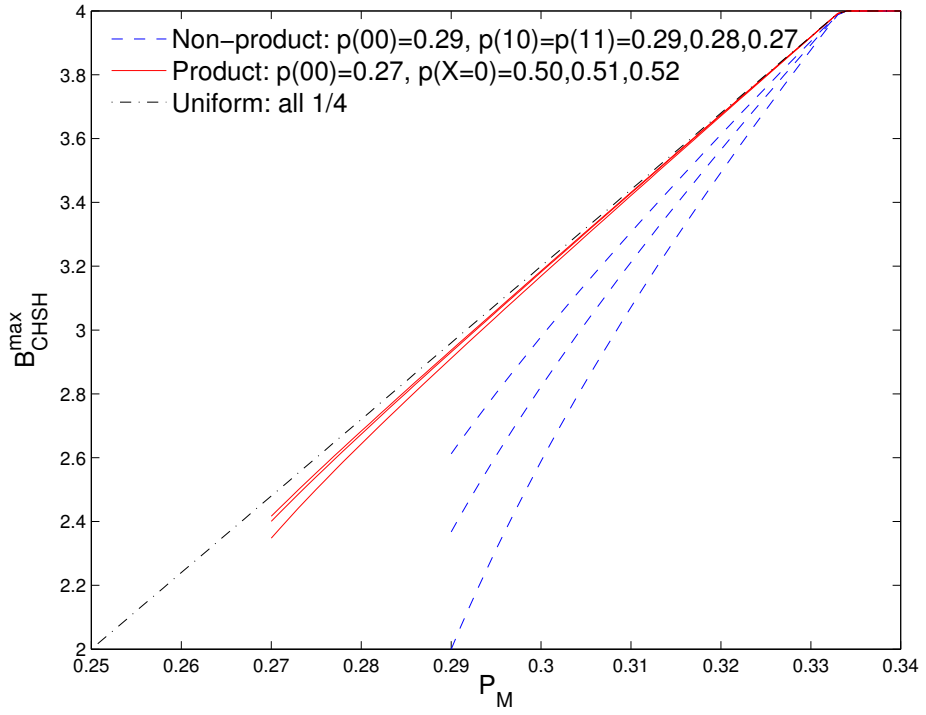

Figure 4.6: Plot of maximum CHSH value against measurement dependence P_M for different p_{obs}. Notice that P_M start from $\max\{p_{00}, p_{01}, p_{10}, p_{11}\}$ because of the data processing inequality: no underlying model of smaller P_M can reproduce the observed input statistics. In a Bell test with an assumed dependence bound P_M, if the value of the inequality is above the line that corresponds to the observed input distribution p_{obs} then there is intrinsic randomness in the outcomes contributed by λ. Therefore, for some observed violations, biased settings statistics can allow the certification of intrinsic randomness while uniform statistics cannot.

for $\max_{x,y} p_{obs}(xy) \leq P_M < \frac{1}{3}$ ($H^{obs}_{\min}(xy) \geq H_{\min}(xy|\lambda) > \log 3$). Notice that when the distribution is uniform $q = 16$ and the second term vanishes, leaving a linear expression in P_M.

It is interesting to observe that for the purpose of violating Bell inequalities (that is, demonstrating non-locality by exceeding $B^{\max}$) under measurement dependence, suppose the inputs have privacy quantified by P_M, then it is advantageous for us to *purposely* select an input distribution that is not uniform. This can be seen easily from for example the red curves for $P_M = 0.27$: selecting uniform input distribution allows a violation up to about 2.5 while selecting non-uniform input distribution only allows a lower maximum violation! Note that for non-uniform distributions on the inputs the upper bound on the Bell value is only as low as 2 (the local bound assuming measurement independence) for non-product distributions on the inputs. (See the blue dashed curves.) All non-uniform product input distributions can have

Bell values larger than 2, if measurement independence is relaxed. Notice also, that the lowest blue curve, the one that takes the value 2 at $P_M = 0.29$ is the one corresponding to the distribution $[p(00), p(01), p(10), p(11)] = [0.29, 0.13, 0.29, 0.29]$. This is precisely the form of the distribution on the inputs we has anticipated at the start of this section.

4.2.4(b) Generalizations

We have seen that for the case of the CHSH inequality, the strategy outlined in Section 4.2.3 in equation (4.51) is the optimal strategy even in the case that the distribution $p_{obs}(xy)$ is not uniform. In general however this is not the case. It is possible to find some inequalities that together with some distributions $p_{obs}(\mathbf{z})$ do not admit a strategy of the form

$$p(xy|\lambda_{\bar{\mathbf{z}}}) = \begin{cases} P_M, & \text{if } \mathbf{z} \in \mathcal{S}_{\bar{\mathbf{z}}} \\ Q(\lambda_{\bar{\mathbf{z}}}), & \text{otherwise} \end{cases} \tag{4.62}$$

where $Q(\lambda)$ is determined by the normalization condition to be $Q(\lambda_{\bar{\mathbf{z}}}) = \frac{1-|S_g^B(\lambda)|}{|S_h^B(\lambda)|}$.

Let us limit our focus to inequalities with symmetries such that $|S_g^B(\lambda)| = |S_g^B|$ and $|S_h^B(\lambda)| = |S_h^B|$ for all λ. In that case, equation (4.41) can be written as a matrix equation, with $p_{obs}(\mathbf{z})$ and $p(\lambda)$ written as vectors and $p(\mathbf{z}|\lambda_{\bar{x},\bar{y}})$ is a matrix whose entries are defined by equation (4.62). If the $p(\mathbf{z}|\lambda)$ matrix is is full-rank, then there is a unique solution for $p(\lambda)$ that is a valid probability distribution. This will always be the case if $|S_g|$ and $|S_h|$ have no common factors.

Examples of cases where the sizes of the sets S_g and S_h have no common factors are any bipartite Bell inequality with terms for all input pairs present and where both parties have the same number of inputs. For these cases, the min-entropy bound of Section 4.2.3 also applies for any non-uniform observed distribution on the inputs.

If, for a given inequality, $|S_g|$ and $|S_h|$ have at least one common prime factor, there may be some choices of distribution p_{obs} for which the strategy (4.62) will not be able to reproduce p_{obs} with any valid distribution $p(\lambda)$. In that case, the optimal strategy may have to be found numerically. For the i.i.d. case, one do this by solving a linear program that is a generalization of the one presented in the previous section.

4.2.5 Conclusions

Bell tests are an essential tool in device-independent approaches. They rely on a set of reasonable assumptions, but some of the assumptions are untestable. In particular, the correlations between source and settings are *strictly unobservable* and therefore the amount of reduction of measurement independence is ultimately an assumption, either on the power of an adversary, on a physical model for the experiment. This study has demonstrated that when relaxing this assumption, the definition used, be it min-entropy or a Santha-Vazirani condition, is critical with respect to what kind of guarantees can be obtained from a Bell test.

There are results [33, 65] showing that with a Santha-Vazirani source assumption arbitrarily weak randomness can be amplified using a protocol that checks for the violation of a Bell inequality. This cannot be accomplished using a min-entropy condition, as we have demonstrated in Section 4.2.3: for sufficiently low min-entropy any inequality can be violated up to its no-signaling bound, using only the classical measurement dependent correlations and in a way that a third party could predict all of the outcomes of the measurements. Even for the protocol in [32] that amplifies bounded randomness (in the Santha-Vazirani definition) using violations of the chained Bell inequality, in order to get perfectly free bits out, the number of inputs for this inequality must go to infinity. As we point out in 4.2.2(a), in this limit the chained Bell inequality is not robust to any relaxation of input randomness if the min-entropy definition is used instead.

Other works in the literature claiming the amplification of arbitrary min-entropy sources must bypass our result in one way or another. For example, the protocol for block sources described in [39] is able to amplify arbitrary (n, k) *block min-entropy source*, which can be seen a combination of a min-entropy source and a Santha-Vazirani source at the block level: each new block of n bit strings is guaranteed to have at least k bit of min-entropy even when conditioned on previous blocks. If one thinks of the min-entropy source as a one shot resource (use it once and it disintegrates), then a block min-entropy source is a reusable resource, promising at least k bits of min-entropy whenever the source is queried. Indeed, by querying the block source M times, we can create a potentially long string with Mk bits of min-entropy (or equivalently boosting the k value of the original source) which exceeds the threshold presented in Section 4.2.3. The only problem is to "hash" the long string down to obtain the inputs for the Bell test without using any randomness. This is obtained by playing many independent Bell tests (multiple devices) each using an input from a hash function and then XOR one party's output across different tests. Using the same idea, the authors also show

the amplification of a single min-entropy source with high enough *rate* (i.e. high enough k/n). We remark that this idea is also independently discovered by Chung, Shi and Wu in their amplification of min-entropy source paper [69].

The bounds on the min-entropy presented in Section 4.2.3 give immediate bounds for any inequality on the amount of input randomness required to draw conclusions about whether the violation of a Bell inequality can give any certification of quantum or non-local behavior. The method demonstrated for the CHSH inequality in Section 4.2.4 demonstrates how to get tight upper bounds for the value a given Bell inequality can take assuming a min-entropy bound for any distribution over the measurement settings. It also shows that there may be advantages to deliberately choosing non-uniform distributions over measurement settings in device independent protocols, depending on what assumptions are being made. In this direction, perhaps the best development since this work has been published is the result of Pütz *et. al* [127]: it is shown that by selecting the Bell inequality wisely (as a function of the presumed input min-entropy) Bell nonlocality can still be observed for any value of input min-entropy above the inequality independent threshold.

Being as the assumption of measurement independence cannot be confirmed, it is important to understand the consequences for device independent protocols when it is relaxed. It is especially interesting that the min-entropy condition, a condition widely adopted in classical security studies [118, 128–130], is has such a different behavior from the Santha-Vazirani condition for these device-testing purposes.

4.3 Randomness in post-selected data

4.3.1 Why post-selection?

The final step in a Bell-based randomness generation protocol is applying a randomness extractor to the outcomes of the Bell tests. Randomness extractors such as two-universal hashing require an independent random seed: the longer the initial string input to the extractor, the longer the needed seed [5] and the computational time to output the result; in fact, it is an active research direction to construct randomness extractor with short seed length [118]. It would thus be helpful in practice to apply the extractor to a string as short as possible.

In photonic implementations of Bell tests, because of the inefficiencies of the source and the detectors, the recorded data is mostly populated with non-detection

[5] The seedlength is typically the logarithm of the length of the input.

events "$\varnothing$" [131, 132]. From physics, we know that these non-detection events carry little or no randomness: for instance, they may be associated to the source not having emitted any pair in a given time. It is thus tempting to simply discard these events before extracting the randomness. But is it possible to do so without opening the detection loophole [133, 134]?

One may ask a similar question in the following extreme situation. Consider a Bell experiment running over the course of two days. During the first day, the setup works perfectly, producing clicks in Alice's and Bob's detectors at every round which are compatible with a maximal violation of the CHSH inequality. At the beginning of the second day, however, the source of entangled particles stops working, so that no detector click is recorded during the whole day. Any experimentalist witnessing such a behavior would certainly treat the data from each day separately, and maybe even choose to neglect the data accumulated during the second day. However, from a device-independent perspective, one need not assume anything about the behavior of the setup, and thus one may try to reach a conclusion by looking only at the overall statistics over the two days.

In fact, it is not clear why doing so would result in a restrictive estimate in this particular situation, because each of the two regimes can be distinguished very clearly in the data: during the first day, only outcomes 0 and 1 were observed, but the second day only gave rise to non-detections $\varnothing$, which can be formally identified as a third outcome. Available techniques [87, 93] guarantee that these 3-outcomes statistics imply a randomness rate of ~ 0.41 bit/run for Alice's outcomes. At the same time, by performing an analysis on each day independently one could clearly certify 1 bit/run for the first day and 0 bit/run for the second one, resulting in a larger average randomness rate of 0.5 bit/run. Is this mismatch due to the presence of a loophole in the analysis considering each day independently? Is the limitation highlighted here intrinsic for all device-independent method which only depends on the overall statistics?

In order to shed some light on these questions, we provide here a method to quantify the randomness present in a subset of events (for instance, double detections). This method takes into account the whole observed statistics (including the non-detection events) and thus does not open the detection loophole.

Our method applies to i.i.d. sampling in the limit of infinitely many measurement rounds. If randomness cannot be certified in this limit, it can also certainly not be certified in the non-i.i.d. finite statistics case. On the other hand, randomness present in this case might still be certifiable in the non-i.i.d. finite statistics case, but this remains to be proven.

After introducing the method in Section 4.3.2, we analyze its consequence

on several physically-motivated models of observed statistics in Section 4.3.4. A glimpse beyond the i.i.d. restriction is given in Section 4.3.5 before the conclusion.

4.3.2 Average randomness in post-selected data

Consider a bipartite Bell experiment where each party uses inputs $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and obtains outputs $a \in \mathcal{A}$, $b \in \mathcal{B}$. In the following, we consider a bipartition of the joint output alphabet $\mathcal{O} = \mathcal{A} \times \mathcal{B}$ into two sets $\mathcal{V}$ and $\mathcal{N}$ [6]. If the outputs observed at a given round $(a, b) \in \mathcal{V}$, we say that the round is valid, and otherwise, if $(a, b) \in \mathcal{N}$, that it is invalid. The list of all inputs and outputs recorded during the run of the Bell experiment constitutes the raw data of the experiment, whereas we refer to the data observed in valid runs only as the post-selected data. Our goal is to estimate how much randomness can be extracted from the post-selected data.

A priori, an adversary trying to guess the post-selected data might not have access to the information about which run turned out to be valid or invalid, since he should not have access to the outputs observed by the parties. For simplicity, however, we'll assume here that the adversary has access to this information. This allows him to know exactly which run he should try to guess and is thus advantageous for him. This assumption might however be problematic in a non-i.i.d. situation (see Section 4.3.5).

4.3.2(a) Guessing probability with post-selection

In general, the conditional probabilities describing the behavior of the experiment at one round $p(ab|xy)$ can be decomposed as

$$p(ab|xy) = \sum_{\lambda} p(\lambda) p_{\lambda}(ab|xy), \tag{4.63}$$

where $p(\lambda)$ is a normalised distribution and $p_{\lambda}(ab|xy)$ are valid conditional probabilities.

Assuming that an adversary interested in guessing the observed outputs is limited to hold only classical information on the users' devices [87, 135], he can at most have access to the variable λ and to the description of the distributions $p_{\lambda}(ab|xy)$ appearing in the decomposition (4.63). In presence of a given decomposition, his optimal strategy thus consists in guessing the pair of outputs (a, b) which

[6] The label $\mathcal{N}$ is highly motivated by the no-detection event occurring in the practical Bell experiment.

maximizes $p_\lambda(ab|xy)$. His average guessing probability on the raw data is then

$$P_{xy} = \sum_\lambda p(\lambda) \max_{ab} p_\lambda(ab|xy). \tag{4.64}$$

This expression can be operationally interpreted as the (optimal) fraction of runs averaged over λ where the adversary correctly guesses the outcome (a, b) given the knowledge of (x, y, λ) when the runs are played with identical devices behaving independently according to $P(ab|xy)$ (i.e. with i.i.d. devices, see Section 4.3.3).

Note that one reason why an adversary might only have access to classical information about the devices is because he is not allowed to interact directly with the quantum systems (in particular he's not the one who produces them) [135]. For simplicity, we consider here such an adversary: someone who is interested in guessing the outcomes observed by the user, but which is only able to do so based on classical information that he could potentially find. To distinguish him from the adversary usually considered in QKD, which is allowed to directly interact with the quantum systems, we will refer to him as Thomas [87].

Now let us change the rule of the guessing game: at each round the user reveals whether he observes valid outcomes or not, and the adversary is only asked to guess the post-selected data. In this situation, the adversary can win only if he correctly guesses an outcome in $\mathcal{V}$. In other words, whenever a run produces $(a, b) \in \mathcal{N}$, Thomas automatically wins that run. We will show that

$$G_{xy} = \frac{1}{p(V|xy)} \sum_\lambda p(\lambda) \max_{ab\in\mathcal{V}} p_\lambda(ab|xy) \tag{4.65}$$

is the guessing probability when conditioned on the outputs being valid and has the operational interpretation: the adversary wins a fraction G_{xy} of the i.i.d. runs producing outcomes in $\mathcal{V}$. Here, V denotes the event "$(a, b) \in \mathcal{V}$", and so the division by $p(V|xy) = \sum_{ab\in\mathcal{V}} p(ab|xy)$ corresponds to the operation of discarding/ignoring all the runs with outcomes in $\mathcal{N}$. Note that in the case $\mathcal{V} = \mathcal{O}$, i.e. in absence of post-selection, G_{xy} reduces to P_{xy}.

Formally deriving this result requires careful application of the concept of conditional probability, which captures the adversary's information in light of additional knowledge. First notice that for λ with $\max_{ab\in\mathcal{V}} p_\lambda(ab|xy) = 0$, the box $p_\lambda(ab|xy)$ can only produce outcomes which will be discarded. Therefore, these λs will not play any role in the guessing probability. In other words, these λs will not be needed to 'explain' the post-selected data. For the remaining λs, we need to obtain the correct probability of them appearing in the post-selected data in terms

of the available information. The probability of the original box producing valid outcomes is $p(V|xy\lambda) = \sum_{ab\in\mathcal{V}} p_\lambda(ab|xy)$; therefore the normalized non-local boxes given V are

$$p(ab|xy\lambda V) = \begin{cases} \frac{p_\lambda(ab|xy)}{p(V|xy\lambda)} & \text{if } ab \in \mathcal{V}, \\ 0 & \text{if } ab \notin \mathcal{V}. \end{cases} \tag{4.66}$$

Likewise, the probability of λ appearing in the decomposition describing the post-selected data needs to be adjusted to (here we are considering a fixed pair of inputs (x, y))

$$p(\lambda|xyV) = \frac{p(V|\lambda)p(\lambda)}{p(V|xy)}. \tag{4.67}$$

Hence, we obtain a model for the post-selected data

$$p(ab|xyV) = \frac{1}{p(V|xy)} \sum_\lambda p(\lambda|xyV)p(ab|xy\lambda V) \tag{4.68}$$

which is consistent with how the users renormalize the probabilities, namely

$$p(ab|xyV) = \begin{cases} \frac{p(ab|xy)}{p(V|xy)} & \text{if } ab \in \mathcal{V}, \\ 0 & \text{if } ab \notin \mathcal{V}. \end{cases} \tag{4.69}$$

We remark that the sum over λs in the previous decomposition, which is a restricted subset of all the original λs, can be extended to all λs in the original decomposition. This decomposition of the post-selected data into a convex mixture of boxes $p(ab|xy\lambda V)$ allows one to write down the expression for the guessing probability for this pair of settings (see (4.64))

$$G_{xy} = \frac{1}{p(V|xy)} \sum_\lambda p(\lambda|xyV) \max_{ab} p(ab|xy\lambda V) \tag{4.70}$$

which gives (4.65) after unwinding the definitions of post-selected boxes.

4.3.2(b) Optimizing the guessing probability

Since the decomposition of $p(ab|xy)$, namely (4.63), is unknown to us, we must perform an optimization to find the maximum guessing probability of Thomas

under such constraints:

$$\begin{aligned} G^*_{xy} = \max \ & \frac{1}{p(V|xy)} \sum_\lambda p(\lambda) \max_{ab\in\mathcal{V}} p_\lambda(ab|xy) \\ \text{s.t.} \ & \sum_\lambda p(\lambda) p_\lambda(ab|xy) = p(ab|xy) \\ & p(\lambda) \geq 0 \text{ and } \sum_\lambda p(\lambda) = 1 \\ & p_\lambda(ab|xy) \in \mathcal{Q} \end{aligned} \tag{4.71}$$

A priori, it is not clear how many λs need to be considered in this program. The following argument shows that this number can always be assumed finite. Partition the set Λ of all λs into finite number of classes for each $ab \in \mathcal{V}$

$$\Lambda_{ab} = \{\lambda : \arg\max_{a'b'\in\mathcal{V}} p_\lambda(a'b'|xy) = ab\} \tag{4.72}$$

each with probability

$$\tilde{p}(\Lambda_{ab}) = \sum_{\lambda\in\Lambda_{ab}} p(\lambda) \tag{4.73}$$

and average box, which is a convex combination of boxes in the class with weight $p(\lambda)/\tilde{p}(\Lambda_{ab})$,

$$\tilde{p}(a'b'|xy\Lambda_{ab}) = \frac{\sum_{\lambda\in\Lambda_{ab}} p(\lambda) p(a'b'|xy\lambda)}{\tilde{p}(\Lambda_{ab})}. \tag{4.74}$$

Notice that this grouping operation does preserve the property of each class, namely

$$\max_{a'b'\in\mathcal{V}} \tilde{p}(a'b'|xy\Lambda_{ab}) = \tilde{p}(ab|xy\Lambda_{ab}), \tag{4.75}$$

as well as all the constraints in (4.71). This allows us to rewrite the optimization program using only finitely many λs:

$$\begin{aligned} \max \ & \frac{1}{p(V|xy)} \sum_{ab\in\mathcal{V}} \tilde{p}(\Lambda_{ab}) \tilde{p}(ab|xy\Lambda_{ab}) \\ \text{s.t.} \ & \sum_{a'b'\in\mathcal{V}} \tilde{p}(\Lambda_{a'b'}) \tilde{p}(ab|xy\Lambda_{a'b'}) = p(ab|xy) \\ & \tilde{p}(\Lambda_{a'b'}) \geq 0 \text{ and } \sum_{a'b'} \tilde{p}(\Lambda_{a'b'}) = 1 \\ & \tilde{p}(ab|xy\Lambda_{a'b'}) \in \mathcal{Q} \end{aligned} \tag{4.76}$$

From now on, we thus work with a finite number of λs.

Note that when reexpressed in terms of Λ_{ab}, the internal maximization in (4.71) disappeared. This allows us to upper-bound the result of this optimization by a semidefinite program by relaxing the last condition $\tilde{p}(a'b'|xy\Lambda_{ab}) \in \mathcal{Q}$ to just ask

that these probabilities belong to some level of the NPA hierarchy [48, 49, 136]. Such bound can then be easily evaluated numerically.

In the case where no outcomes are discarded, i.e. $\mathcal{V} = \mathcal{O}$, this SDP reduces to the one described in [87, 93] to bound the guessing probability as a function of correlations $p(ab|xy)$.

Notice that the constraints in the above program are based on the distributions *before* post-selection. This reflects the fact that our analysis is not subject to the detection loophole.

4.3.3 A digression: bound for i.i.d. experiments

Here, we discuss the operational interpretation of (4.64) and (4.65), as well as the implication of revealing or not to the adversary the list of valid measurement rounds. As a short hand, z stands for the pair of inputs x, y and c stands for the pair of outputs a, b.

4.3.3(a) The guessing probability for an N-run experiment

Let $\boldsymbol{z}$ and $\boldsymbol{c}$ be the strings of inputs $z_1...z_N$ and outputs $c_1...c_N$ of the Bell experiment. The adversary Thomas holds some additional information $\boldsymbol{\lambda}$ which may be correlated to $\boldsymbol{z}$ and $\boldsymbol{c}$ as characterized by the distribution

$$p(\boldsymbol{c}\boldsymbol{z}\boldsymbol{\lambda}) = p(\boldsymbol{z})p(\boldsymbol{\lambda})p_{\boldsymbol{\lambda}}(\boldsymbol{c}|\boldsymbol{z}), \tag{4.77}$$

where measurement independence is implicitly assumed as the input $\boldsymbol{z}$ and $\boldsymbol{\lambda}$ are independent, and $p_{\boldsymbol{\lambda}}(\boldsymbol{c}|\boldsymbol{z})$ can be seen as the behavior of the experiment or as its "preparation". If the device employs a deterministic strategy then there is only a single $\boldsymbol{\lambda}$ and a single behavior for any experiment; however, if the strategy is probabilistic (as in the case of the following i.i.d. strategy) then different behaviors, i.e. different $p_{\boldsymbol{\lambda}}(\boldsymbol{c}|\boldsymbol{z})$, are prepared in different experiments.

The post-processing we apply to the raw string $\boldsymbol{c}$ can be paraphrased as follows: discard all the symbols c_i in $\mathcal{N}$ and keeping the order of the raw string to get a new string labeled $\boldsymbol{s}$. This can be formalized as a function f from $\mathcal{O}^N$ to $\cup_{m=0}^{N}\mathcal{V}^m$ where $\mathcal{V}^0$ consists of only the empty string (i.e. the raw string $\boldsymbol{c}$ is completely discarded) and $\mathcal{V}^m$ is the set of strings of length m over the alphabet $\mathcal{V}$ (the m times Cartesian product of the set $\mathcal{V}$). Given a post-processed string $\boldsymbol{s}$, we denote $f^{-1}(\boldsymbol{s})$ as the set of raw strings which are mapped to $\boldsymbol{s}$ under f. Thomas can "undo" the effect of our post-processing by considering the induced distribution of

the post-selected data

$$q_{\boldsymbol{\lambda}}(\boldsymbol{s}|\boldsymbol{z}) = \sum_{\boldsymbol{c}\in f^{-1}(\boldsymbol{s})} p_{\boldsymbol{\lambda}}(\boldsymbol{c}|\boldsymbol{z}). \tag{4.78}$$

Thomas' guessing probability given his available information gives us a bound on how many uniformly random bits we can extract by hashing the post-processed string.

The description so far is quite general; it encompasses devices behaving in a non-i.i.d. manner. The restriction to the i.i.d. behavior, where in each run a box labeled $p_{\lambda_i}(c_i|z_i)$ is sampled from a fixed set of boxes with probability $p(\lambda_i)$, simplifies the form of $p_{\boldsymbol{\lambda}}(\boldsymbol{c}|\boldsymbol{z})$ greatly. In other words, the behavior

$$p_{\boldsymbol{\lambda}}(\boldsymbol{c}|\boldsymbol{z}) = p_{\lambda_1}(c_1|z_1)...p_{\lambda_N}(c_N|z_N) \tag{4.79}$$

is prepared in an experiment with probability $p(\boldsymbol{\lambda}) = p(\lambda_1)...p(\lambda_N)$. The reason this is called i.i.d. is because when average over many experiments

$$p(\boldsymbol{c}|\boldsymbol{z}) = \sum_{\lambda} p(\lambda)p_{\lambda}(c|z) \tag{4.80}$$

$$= \prod_{i=1}^{N}\left(\sum_{\lambda_i} p(\lambda_i)p_{\lambda_i}(c_i|z_i)\right) = \prod_{i=1}^{N}\hat{p}(c_i|z_i) \tag{4.81}$$

where $\hat{p}(c_i|z_i)$ is the single run average behavior.

The sum in (4.78) can be decomposed as follows: the *structure of a string* $\boldsymbol{c}$ is given by two set of indices α_i for which $c_{\alpha_i} \in \mathcal{V}$ and β_j for which $c_{\beta_j} \in \mathcal{N}$. For instance, the string $\boldsymbol{c} = 0\varnothing10\varnothing$ has structure $\{1,3,4\}$ as indices in valid and $\{2,5\}$ as indices in no-detection (i.e. the structure is $\mathcal{V}\mathcal{N}\mathcal{V}\mathcal{V}\mathcal{N}$). Given $\boldsymbol{s}$ with length m then the set $f^{-1}(\boldsymbol{s})$ can be grouped into strings with the same structure, allowing us to rewrite

$$q_{\boldsymbol{\lambda}}(\boldsymbol{s}|\boldsymbol{z}) = \sum \sum_{c_{\beta_j}\in\mathcal{D}} \prod_{i=1}^{m} p_{\lambda_{\alpha_i}}(s_{\alpha_i}|z_{\alpha_i}) \prod_{j=1}^{N-m} p_{\lambda_{\beta_j}}(c_{\beta_j}|z_{\beta_j}) \tag{4.82}$$

$$= \sum \prod_{i=1}^{m} p_{\lambda_{\alpha_i}}(s_{\alpha_i}|z_{\alpha_i}) \prod_{j=1}^{N-m} p_{\lambda_{\beta_j}}(\mathcal{N}|z_{\beta_j}) \tag{4.83}$$

for $\boldsymbol{s} \in \mathcal{V}^m$, where the outermost sum is over structures compatible with $\boldsymbol{s}$, and $p_{\lambda_{\beta_j}}(\mathcal{N}|z_{\beta_j}) = \sum_{c_{\beta_j}\in\mathcal{N}} p_{\lambda_{\beta_j}}(c_{\beta_j}|z_{\beta_j})$ is the probability of no-detection for the β_j^{th} run.

Now the string guessing probability for three different scenarios can be discussed. If we reveal nothing to Thomas other than the specified protocol, then Thomas' optimal strategy is to pick the most probable post-processed string according to

$q_{\boldsymbol{\lambda}}(\boldsymbol{s}|\boldsymbol{z})$, i.e. we have the guessing probability

$$P_{guess}(\boldsymbol{S}|\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\Lambda}=\boldsymbol{\lambda}) = \max_{\boldsymbol{s}} q_{\boldsymbol{\lambda}}(\boldsymbol{s}|\boldsymbol{z}) \tag{4.84}$$

for the given experiment. If, however, we reveal to Thomas the length of the post-processed string, then Thomas can restrict his search of the most probable string to the set $\mathcal{V}^m$ thus giving

$$P^m_{guess}(\boldsymbol{S}|\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\Lambda}=\boldsymbol{\lambda}) = \max_{\boldsymbol{s}\in\mathcal{V}^m} \frac{q_{\boldsymbol{\lambda}}(\boldsymbol{s}|\boldsymbol{z})}{\sum_{\boldsymbol{s}\in\mathcal{V}^m} q_{\boldsymbol{\lambda}}(\boldsymbol{s}|\boldsymbol{z})}. \tag{4.85}$$

Finally, if we even reveal to Thomas the structure of the post-processed string then he only need to guess among the strings with that structure:

$$P^{\text{valid}}_{guess}(\boldsymbol{S}|\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\Lambda}=\boldsymbol{\lambda}) = \prod_{i=1}^{m} \max_{s_{\alpha_i}\in\mathcal{V}} \frac{p_{\lambda_{\alpha_i}}(s_{\alpha_i}|z_{\alpha_i})}{p_{\lambda_{\alpha_i}}(\mathcal{V}|z_{\alpha_i})}. \tag{4.86}$$

On average over many realizations (with the same input string $\boldsymbol{z}$ and post-selected length m),

$$\begin{aligned} P^{\text{valid}}_{guess}(\boldsymbol{S}|\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\Lambda}) &= \sum_{\boldsymbol{\lambda}} p(\boldsymbol{\lambda}) P^{\text{valid}}_{guess}(\boldsymbol{S}|\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\Lambda}=\boldsymbol{\lambda}) \\ &= \prod_{i=1}^{m} \left[\sum_{\lambda_{\alpha_i}} p(\lambda_{\alpha_i}) \max_{s_{\alpha_i}\in\mathcal{V}} \frac{p_{\lambda_{\alpha_i}}(s_{\alpha_i}|z_{\alpha_i})}{p_{\lambda_{\alpha_i}}(\mathcal{V}|z_{\alpha_i})} \right] = \prod_{i=1}^{m} G_{z_{\alpha_i}} \end{aligned}$$

where $G_{z_{\alpha_i}}$ can be seen as the single run average guessing probability with post-selection.

Therefore, we are led to bound the *single run* guessing probability

$$\max_{s_{\alpha_i}\in\mathcal{V}} \frac{p_{\lambda_{\alpha_i}}(s_{\alpha_i}|z_{\alpha_i})}{p_{\lambda_{\alpha_i}}(\mathcal{V}|z_{\alpha_i})} \tag{4.87}$$

and its average over λ, namely

$$\sum_{\lambda_{\alpha_i}} p(\lambda_{\alpha_i}) \max_{s_{\alpha_i}\in\mathcal{V}} \frac{p_{\lambda_{\alpha_i}}(s_{\alpha_i}|z_{\alpha_i})}{p_{\lambda_{\alpha_i}}(\mathcal{V}|z_{\alpha_i})}, \tag{4.88}$$

as presented in Section 4.3.2.

4.3.4 Examples relevant for experiments

Let us consider an application of the guessing game described in the previous section on the correlations that can be sampled in some quantum experiments. For

the sake of the example, we consider a simplified model of the correlations expected in a pulsed down-conversion experiment. The source is an on-demand photon pair source with finite efficiency ν: a two-photon state is produced with probability ν and the vacuum with probability $1-\nu$. Each measurement setup of each party has two detectors, which might have a finite efficiency η (this efficiency also takes into account possible losses along the transmission of the photon). Because of both the vacuum component and the inefficiency of the detectors, a third outcome is possible in each run, namely the absence of any detection. The fourth possible outcome that could appear in a real experiment, double detection, is neglected; and so are dark counts.

Note that it is common in photonic experiments to use a single detector to attribute the result 0 or 1 to a measurement. This is achieved by assigning one result in case of a detection, and the other one in absence of detection. In doing so, no distinction is made between the absence of detection which is due to loss and that which comes from the state of the photon. For simplicity, we don't consider this situation here. It can be analysed similarly, though, and one can check that it provides similar conclusions for the scenarios considered here.

In our situation, each party's measurement has three possible outcomes called $0, 1, \varnothing$. If $p(ab|xy)$ for $a, b, x, y \in \{0, 1\}$ is the correlation obtained from some choice of state and measurements when $\eta = \nu = 1$, then in the presence of imperfect source and detectors the observed statistics are

$$q(ab|xy) = \begin{cases} \nu\eta^2 p(ab|xy) & \text{if } a, b \in \{0,1\}, \\ \nu\eta(1-\eta)p(a|x) & \text{if } a \in \{0,1\} \text{ and } b = \varnothing \\ \nu\eta(1-\eta)p(b|y) & \text{if } a = \varnothing \text{ and } b = \{0,1\} \\ \nu(1-\eta)^2 + (1-\nu) & \text{if } a = b = \varnothing, \end{cases} . \tag{4.89}$$

where $p(a|x)$ and $p(b|y)$ are marginal distributions of $p(ab|xy)$.

Using the program (4.76), we are going to compute lower bounds on the extractable randomness that can be found in the following cases:

- (a) All outcomes are considered in the SDP program, i.e. $\mathcal{N} = \mathcal{N}_{\text{a}} = \emptyset$ (the empty set).
- (b) The post-selected string of outcomes does not contain double occurrences of $\varnothing$, i.e. $\mathcal{N} = \mathcal{N}_{\text{b}} = \{\varnothing\varnothing\}$.
- (c) The post-selected string of outcomes does not contain any occurrence of a no-detection event $\varnothing$, i.e. $\mathcal{N} = \mathcal{N}_{\text{c}} = \{0\varnothing,\ 1\varnothing,\ \varnothing\varnothing,\ \varnothing 0,\ \varnothing 1\}$.

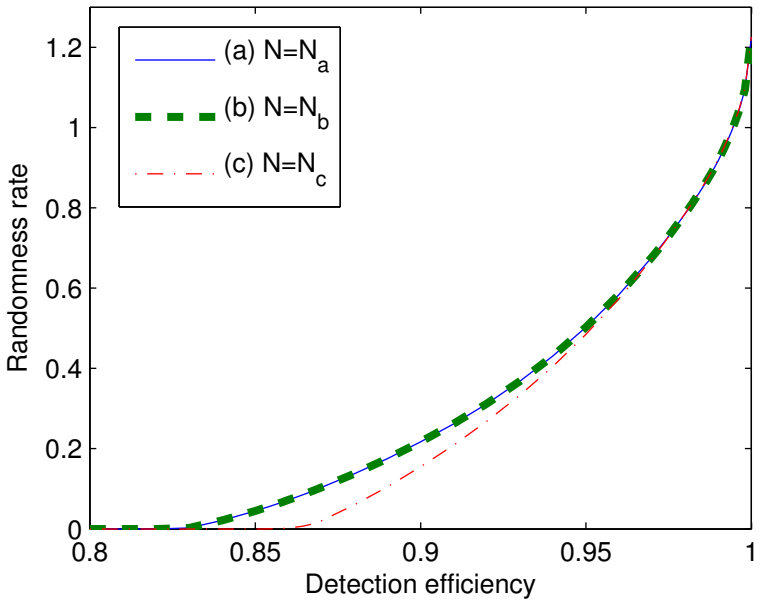

Figure 4.7: Randomness from a singlet with finite detection efficiency. Curves (a) and (b) coincide almost perfectly and approach 0 at the detection loophole limit 0.828 [134].

Removing some outcomes reduces the length of the considered data by different amounts depending on which post-selection process (a), (b) or (c) is used. Hence, the randomness rate $-\log_2 G^*_{xy}$ corresponds to different length of the input string to the randomness extractor. For a fair comparison, we renormalize the randomness rate obtained on post-selected data with respect to the total number of runs. This choice is also natural because the total number of runs is the actual amount of resources used in the randomness generation process.

Finally, let us mention that all optimizations reported here were performed at local level 1 of the SDP hierarchy [137].

4.3.4(a) Example 1: Imperfect detectors, perfectly heralded source

We first consider the case where the source efficiency is $\nu = 1$ and η varies. This is the model that was used in most studies of the detection loophole, even if it is not realistic for down-conversion, because one cannot suppress the vacuum component without producing higher-number excitations (the joint study of detection loophole and realistic models of the source in down-conversion experiments is surprisingly recent [138, 139]).

For the state, we consider two cases. First, that of a maximally-entangled state $|\psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. In this case, the usual optimal CHSH measurements for

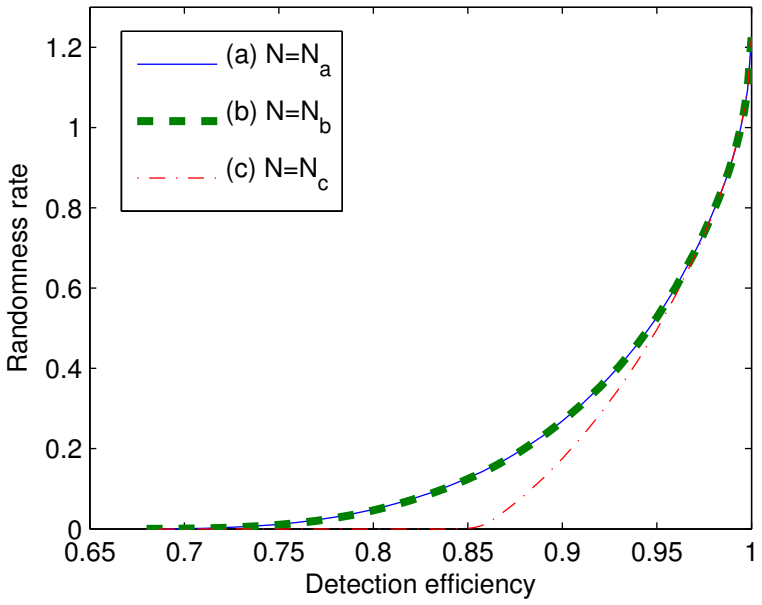

Figure 4.8: Randomness from Eberhard correlations. Curves (a) and (b) coincide and approach 0 at the Eberhard limit of 2/3 [133].

the singlet give

$$p(ab|xy) = \frac{1}{4}\left(1 + (-1)^{a+b+xy}\frac{1}{\sqrt{2}}\right) \tag{4.90}$$

for $a, b, x, y \in \{0, 1\}$. The expected randomness rate as a function of η is shown in Figure 4.7. We note that no randomness can be extracted if $\eta \leq 82.8\%$ which is the boundary at which those statistics can be explained locally with a model exploiting the detection loophole.

The second case is that of Eberhard's famous study [133]. The state is the partially entangled state $|\psi\rangle = \cos(\theta)\,|00\rangle + \sin(\theta)\,|11\rangle$ with θ depending on the detector efficiency η; in turn, the measurement can be parametrized by two angles α_0, α_1 which also depend on η. These parameters are chosen to optimize the violation of a lifting [140] of the CHSH inequality, for each value of η. The resulting randomness rate is plotted in Figure 4.8. Again, no randomness can be extracted below the detection loophole threshold of $\eta \leq 66.6\%$.

In both cases, we notice that about the same amount of randomness is certified in case (a) and (b) (with a small difference of less than 10% in advantage of case (a) for Eberhard correlations when the efficiency is $\eta \,\tilde{\in}\, [0.7, 0.9]$). Extracting randomness from the subset of all the data in which the double non-detection events $(\varnothing, \varnothing)$ are removed thus doesn't seem to be a problem here: in a heralded Bell test with finite detection efficiency, one can essentially discard double no-detection events, provided that a count is kept of how many of them appear (their statistics

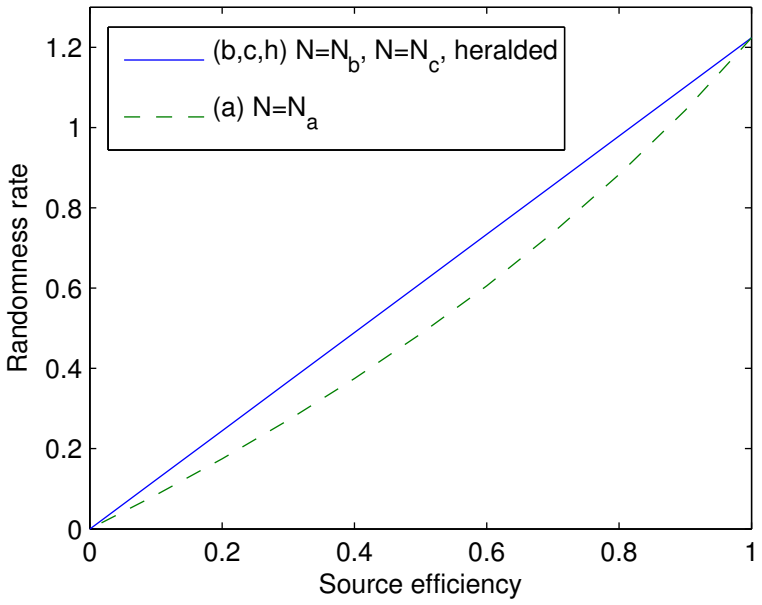

Figure 4.9: Randomness from a singlet produced with finite probability. Curves (b) and (c) are identical, since there are no events with one detection and one no-detection in the raw data (the post-selection procedures (b) and (c) are actually the same for this correlation). Curve (h), which gives the randomness from raw string of outcomes upon the heralding of a successful preparation of the state (i.e. randomness from the correlation (4.90)), exactly coincides with curves (b) and (c). Curve (a) lies below the other ones.

is kept).

However, considering case (c), we see that removing all events where some no-detection occurred results in a clearly lower randomness rate. For efficiencies lower than 86% and 85%, no randomness at all is even certified. This kind of post-selection is thus too strong if one is interested in certifying an optimal amount of randomness.

4.3.4(b) Example 2: Perfect detectors, not heralded source

The other limiting case, where the detectors have perfect efficiency but not the source, has not received much attention so far. However, this is going to be the ideal case for experiments using down-conversion sources: coupling and detection efficiencies are being steadily improved, but one cannot avoid the fact that down-conversion produces almost perfect singlets only in the regime $\nu \ll 1$.

In Figure 4.9, we show how much randomness can be certified when $q(ab|xy)$ is given in (4.90), detection efficiency is $\eta = 1$ and ν varies. In this case, the lower bound on the randomness computed from the raw data is *lower* than the one obtained after removing double no-detections from the data. In fact, after

discarding double no-detections, the same amount of randomness that could be certified if the source was heralded is recovered (i.e. it is proportional to the source efficiency ν).

That the bounds we find are never better than the heralded one is consistent with the intuition that double non-detections contain no randomness. However, the fact that *more* randomness can be certified from considering the post-selected data than from considering the raw data may sound surprising and even potentially wrong: did some additional randomness come from the post-processing? Let us show that this result is indeed correct.

For this, we go back to the optimization program (4.71). First absorbing $p(\lambda)$ in to $p_\lambda(ab|xy)$ and calling the new variable $P_\lambda(ab|xy)$, the optimization for randomness in setting pair (x, y) of the post-selected data can be rewritten as

$$\begin{aligned} \max \ & \frac{1}{p(V|xy)} \sum_\lambda \max_{ab\in\mathcal{V}} P_\lambda(ab|xy) \\ \text{s.t.} \ & \sum_\lambda P_\lambda(ab|xy) = p(ab|xy) \ \forall abxy \\ & P_\lambda(ab|xy) \in \mathcal{Q} \text{ sub-normalized} \end{aligned}$$

where $\mathcal{V} = \{00, 01, 10, 11, 0\varnothing, 1\varnothing, \varnothing 0, \varnothing 1\}$ if we are going to discard double non-detections. Now, for the case under study, we have

$$p(ab|xy) = \begin{array}{cc||ccc|ccc} & \mathbf{x} & & \mathbf{0} & & & \mathbf{1} & \\ \mathbf{y} & b\backslash a & 0 & 1 & \varnothing & 0 & 1 & \varnothing \\ \hline & 0 & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \mathbf{0} & 1 & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ & \varnothing & 0 & 0 & 1-\nu & 0 & 0 & 1-\nu \\ \hline & 0 & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \mathbf{1} & 1 & \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ & \varnothing & 0 & 0 & 1-\nu & 0 & 0 & 1-\nu \end{array} \tag{4.91}$$

where the sum of the probabilities in the dotted positions in each sector sums up to ν. This means that, for all λ, the underlying strategy must have zeros in the

positions below:

$$P_\lambda(ab|xy) = \left(\begin{array}{ccc|ccc} \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & 0 & 0 & \cdot \\ \hline \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & 0 & 0 & \cdot \end{array}\right). \tag{4.92}$$

In turn therefore, any underlying strategy with a guessing probability $G_{xy}(\{P_\lambda(ab|xy)\})$ can be reduced to an underlying strategies of the form

$$Q_\lambda(ab|xy) = \left(\begin{array}{ccc|ccc} \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \ \lambda \neq (\varnothing, \varnothing) \tag{4.93}$$

$$Q_{\lambda=(\varnothing,\varnothing)}(ab|xy) = \left(\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\nu & 0 & 0 & 1-\nu \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\nu & 0 & 0 & 1-\nu \end{array}\right), \tag{4.94}$$

(which utilizes one additional lambda as compared to the original strategy) while still satisfying $G_{xy}(\{Q_\lambda(ab|xy)\}) = G_{xy}(\{P_\lambda(ab|xy)\})$. The optimization for randomness can be further simplified to

$$\begin{aligned} \max \ & \frac{1}{\nu} \sum_{\lambda \neq \varnothing\varnothing} \max_{ab \in \mathcal{V}} Q_\lambda(ab|xy) \\ \text{s.t.} \ & \sum_{\lambda \neq \varnothing\varnothing} Q_\lambda(ab|xy) = p(ab|xy) \ \forall ab \in \mathcal{V} \ \forall xy \\ & Q_\lambda(ab|xy) \in \mathcal{Q} \text{ sub-normalized} \end{aligned}$$

where we have emphasized the removal of the strategy $\lambda = (\varnothing, \varnothing)$ in the objective function as well as in the constraints. It is easily seen then that this optimization is the same as that of a heralded source where the observed correlation is $p_h(ab|xy) =$

$p(ab|xy)/\nu$ (namely the following optimization

$$\begin{aligned} \max \ & \sum_{\lambda} \max_{ab} q_\lambda(ab|xy) \\ \text{s.t.} \ & \sum_{\lambda} q_\lambda(ab|xy) = p_h(ab|xy) \ \forall abxy \\ & q_\lambda(ab|xy) \in \mathcal{Q} \text{ sub-normalized} \end{aligned}$$

with $q_\lambda(ab|xy) = Q_\lambda(ab|xy)/\nu$ — observe that in this last optimization all probabilities are with respect to the heralded data or post-selected data, unlike the previous optimization where probabilities are computed on the raw data including the $(\varnothing, \varnothing)$ outcome). This gives curve (h) of Figure 4.9.

In other words, the mechanism for the increase in min-entropy is a consequence of the properties of the min-entropy with respect to the state: it is neither a concave or a convex function of the state. Consider the toy example of mixing two coins where the first is a perfect coin outputting 0 or 1 with min-entropy 1 and the second is a deterministic coin outputting 2 or 3 with min-entropy 0 for instance. If the mixing is uniform, namely each coin is chosen with probability 1/2, then we have a four sided die with outcome alphabet 0,1,2,3 and min-entropy 0.415. If we flip only once the two coins, we get 1 bit of randomness for the two outcomes. If we flip the four sided die we get only a single output among 0,1,2,3 and this outcome has 0.415 bit of randomness. Now we are trying to reverse direction of this mixing, and the fact that the min-entropy is neither concave nor convex, gives us no limitation on the amount of randomness we can gain. In fact, for this example we gain from 0.415 bit (alphabet 0,1,2,3) to a full 1 bit (alphabet 0,1). (To be fair, this happen only half of the time because the other half we get the second coin with output alphabet 2 or 3 which has no randomness). In other words, we managed to decompose the state from the four sided die to two coins with "additional information" that the process is actually mixed as described. Likewise, in our previous analysis with the SDP, the "additional information" is provided by the form of the global correlation.

Also, there is no contradiction with the data processing inequality because the post-processing can only increase the min-entropy. The min-entropy of post-selected data can be lower bounded by the min-entropy in the raw data, which in turn can be provided by the method in [87]. Here our analysis lower bounded the min-entropy of post-selected data directly using the observed correlations, and is able to obtain a better bound.

This also resolves the two-day paradox mentioned in the introduction (at least in its i.i.d. randomized form). Indeed, doing a similar analysis as shown in Figure 4.9

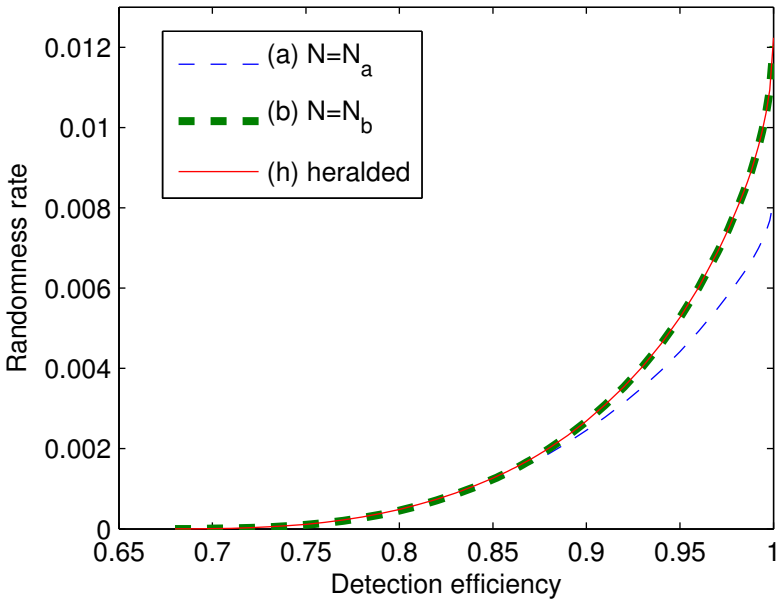

Figure 4.10: Randomness from the Eberhard correlations produced with 1% probability. Curve (a) is lower than if the source was heralded (curve (h)). However, curve (b) recovers almost perfectly the heralded case (h).

for the randomness of Alice's outcomes confirms that the average rate of 1/2 bit per run can be certified by only considering the global statistics $P(ab|xy)$.

This reasoning shows that the optimal way of certifying randomness may be achieved with some processing applied before the extraction (hence a pre-processing as seen from the point of view of randomness extraction). A similar observation was made for QKD, where adding random local noise may actually help [20]. In the case we are studying, one may see this pre-processing as dividing the observed outcomes into two strings, one that contains $(00, 01, 10, 11)$ and the other that contains $(\varnothing\varnothing)$. The randomness is then obtained by applying an extractor on both strings (extraction on the second string being trivial in this case), and concatenating the resulting uniform strings [7]. Clearly, such an analysis by part could be applied for any partitioning of the joint set of outputs $\mathcal{O}$, as long as the extraction parameters for each part takes into account the original statistics on the set of all outputs.

4.3.4(c) Example 3: Imperfect detectors, not perfectly heralded source

Finally we study an example where both the detectors and the source are imperfect. Now, no-detection events can come either from the source having left the field in

[7] One must be careful with the resulting quality (as measured by distance to uniform) of the combined string.

the vacuum state, or from the detectors not having fired. In Figure 4.10, we set $\nu = 1\%$ and vary η for the Eberhard correlations described in 4.3.4(a).

Again we find that slightly more randomness can be certified by considering $\mathcal{N} = \mathcal{N}_{\mathrm{a}}$ when the efficiency is low ($\eta \lesssim 0.85$). However, the analysis with $\mathcal{N} = \mathcal{N}_{\mathrm{b}}$ is advantageous whenever $\eta \gtrsim 0.85$. Moreover, it gives a similar bound as curve (a) from Figure 4.8 rescaled by a factor of ν. Thus, roughly as much randomness as could be certified if the source was heralded could also be extracted here, by applying an extractor on just less than 1% of the raw data.

4.3.5 Beyond the i.i.d. case

In this section, we are going to show that relaxing the i.i.d. assumption in the scenario considered above results in strictly less randomness being certifiable than shown above. The fact that non-i.i.d. strategies are strictly more powerful than i.i.d. strategies *even in the asymptotic limit of infinitely many runs* is, to our knowledge, a feature not found in previous works on randomness from Bell tests [111, 115, 141] or on quantum key distribution [23].

While we are not able to provide a general bound against non-i.i.d. devices, we can provide explicit examples of more powerful strategies. These non-i.i.d. strategies exploit the fact that we are revealing whether the outcomes of a run are discarded or not. They can thus be easily circumvented by not letting the user reveal any information about the outputs that he observes, thus potentially reconciling the asymptotic i.i.d. and non-i.i.d. bounds. As mentioned in Section 4.3.2, such scenario has other complications that prevented its study so far.

Specifically, we found the strategies below by thinking in the following terms: suppose that the outcomes of run n have been kept; the adversary would like to know the value. But the boxes may happen to behave in such a way that the fact of keeping or discarding the outcome at run $n+1$, which Thomas will learn, leaks some information about the outcome kept at run n. This is similar to the argument of [74].

For clarification, let us mention that the kind of non-i.i.d. behavior considered here is possible even in a situation in which the adversary does not produce the devices. Simply, it is not a consequence of the action of a possibly malicious adversary, but rather represent a possible defect of the devices. If such defects are compatible with the observed statistics, one cannot exclude a priori that a non-malicious adversary could become aware of it, and exploit it in his guessing strategy.

4.3.5(a) An explicit strategy

Consider the following quantum correlations, as written in terms of Collins-Gisin tables [142]:

$$P_1 = \begin{array}{|c|c|c|} \hline 1 & 0.4453 & 0.3121 \\ \hline 0.6570 & 0.1708 & 0.0394 \\ 0 & 0 & 0 \\ \hline 0.3244 & 0.0247 & 0.2843 \\ 0.4942 & 0.4195 & 0.0277 \\ \hline \end{array}, \tag{4.95}$$

$$P_2 = \begin{array}{|c|c|c|} \hline 1 & 0.8544 & 0.7373 \\ \hline 0 & 0 & 0 \\ 0.8919 & 0.8381 & 0.7209 \\ \hline 0.2619 & 0.1165 & 0.2617 \\ 0.4973 & 0.4972 & 0.2354 \\ \hline \end{array}, \tag{4.96}$$

$$P_3 = \begin{array}{|c|c|c|} \hline 1 & 0.6042 & 0.5429 \\ \hline 0.3979 & 0.0886 & 0.0365 \\ 0.6021 & 0.5156 & 0.5064 \\ \hline 0.4588 & 0.1078 & 0.4267 \\ 0.5412 & 0.4964 & 0.1162 \\ \hline \end{array}, \tag{4.97}$$

$$P_4 = \begin{array}{|c|c|c|} \hline 1 & 0.6663 & 0.2038 \\ \hline 1 & 0.6663 & 0.2038 \\ 0 & 0 & 0 \\ \hline 0.2936 & 0.1393 & 0.1112 \\ 0.7064 & 0.5270 & 0.0926 \\ \hline \end{array}, \tag{4.98}$$

$$P_5 = \begin{array}{|c|c|c|} \hline 1 & 0.9996 & 0.0015 \\ \hline 0 & 0 & 0 \\ 1 & 0.9996 & 0.0015 \\ \hline 0.0010 & 0.0006 & 0.0004 \\ 0.9990 & 0.9990 & 0.0011 \\ \hline \end{array}. \tag{4.99}$$

Now consider a device which realises one of the three first correlations with probability $p(1) = 0.4097$, $p(2) = 0.4992$, $p(3) = 0.0911$. Whenever one of the first two boxes is chosen, it determines the outcomes for that run, and a new box within this same set is chosen for the next round. When the third box is chosen, it also determines the outcomes for that run, but the outcome of the next round are sampled as follows:

- If Alice's outcome in the current run is 1, and the value of two pre-determined hidden variables λ and μ are 0, then the box P_4 is chosen for the next round.
- If $a = 1$, $\lambda = 0$, $\mu = 1$, then Bob still uses his part of box P_4, but Alice outputs the third outcome.
- If $a = 1$, $\lambda = 1$, $\mu = 0$, then box P_5 is chosen for the next round.
- If $a = 1$, $\lambda = 1$, $\mu = 1$, then Bob uses P_5 and Alice outputs the third outcome.
- If $a = 2$, $\lambda = 0$, $\mu = 0$, then Bob uses P_4 and Alice outputs the third outcome.
- If $a = 2$, $\lambda = 0$, $\mu = 1$, then box P_4 is chosen for the next round.
- If $a = 2$, $\lambda = 1$, $\mu = 0$, then Bob uses box P_5 and Alice outputs the third outcome.
- If $a = 2$, $\lambda = 1$, $\mu = 1$, then box P_5 is chosen for the next round.
- (Note that box P_3 never produces $a = 3$.)

Here μ and λ are i.i.d. variables with distribution $p(\lambda = 0) = 1 - p(\lambda = 1) = 0.0013$, $p(\mu = 0) = p(\mu = 1) = 1/2$, and different realizations of λ, μ are used every time the box number 3 is chosen.

Thanks to the fact that boxes P_4 and P_5 only produce one of the first two possible outcomes, Alice's outcome when box 3 is used is always *fully encoded* in the fact that his outcome in the next round be in the valid $\mathcal{V}$ or invalid $\mathcal{N}$ set, and in the knowledge of variable μ for that run. Moreover, for all boxes except P_3 one of the first two outcomes always appears with probability zero. It is thus always possible to guess Alice's outcome.

One can check, however, that it would not be possible to fully guess Alice's outome if the device behaved in an i.i.d manner. For this, note that the average observed correlations according to the above rules are $P =$

$$\frac{p(1)P_1 + p(2)P_2 + p(3)P_3 + p(3)(p(\lambda = 0)(P_4 + P_4^B)/2 + p(\lambda = 1)(P_5 + P_5^B)/2)}{p(1) + p(2) + 2p(3)} \tag{4.100}$$

which numerically evaluates to

$$\begin{array}{c|c|c} 1 & 0.6919 & 0.5000 \\ \hline 0.2800 & 0.0716 & 0.0178 \\ 0.5000 & 0.4681 & 0.3722 \\ \hline 0.2800 & 0.0716 & 0.2621 \\ 0.5000 & 0.4681 & 0.1279 \end{array} , \tag{4.101}$$

where P_4^B and P_5^B denote the correlations obtained when Bob uses the box 4 or 5, and Alice produces a third outcome.

Applying our i.i.d. SDP to these correlations, one can show that in case Alice uses her first input and the run is not discarded, the guessing probability on her outcome is upper-bounded by 0.9874.

4.3.5(b) A simpler but less realistic strategy

The above strategy requires information about outcomes observed in previous runs, but never from a different box: neither Alice's nor Bob's device needs to know which outcomes the other party observed in any previous run. If one were to allow Alice and Bob's devices to depend on all of their previous inputs and outputs, a simpler strategy could already be possible. Note however that one would have to argue why it is the case that one box can signal to another one, but not to the adversary!

Consider the situation in which the parties measure a singlet with probability p, and nothing with probability $1-p$. This situation doesn't create any single no-detection: no-detections always come in pair between Alice and Bob. Thus, we know that for any $p > 0$, some randomness remains in the non-discarded outcomes (see Figure 4.9).

However, the same statistics could be observed if measurements are always performed on a perfect singlet, and runs with double no-detections are artificially added by using the following rule:

singlet outcomes a, b	following runs
$1, 1$	
$1, 2$	$\varnothing, \varnothing$
$2, 1$	$\varnothing, \varnothing$ $\varnothing, \varnothing$
$2, 2$	$\varnothing, \varnothing$ $\varnothing, \varnothing$ $\varnothing, \varnothing$

(4.102)

In this case, counting the number of successive discarded events fully informs about the value of both parties' outcomes. This corresponds to the above situation with an average source efficiency of $p = 2/5 > 0$.

4.3.6 Conclusions

In this study we consider the effect of post-selection on Bell based randomness generation. We showed that it is possible to analyse the randomness that can be extracted from post-selected data without opening the detection loophole.

For several physically-motivated models of the observed statistics, we showed that one can indeed vindicate the idea that most of the randomness is present in the double-detection events. From a practical perspective, this means that we can directly hash the post-selected subset of the original data, which is of much smaller size for current efficiencies, thereby reducing the needed seed length, and also the time required to compute the final output. In the case of the statistics created in a down-conversion experiment with low pumping, in particular, our result suggest that essentially all the randomness can be extracted from the set of data where at least one party observed a detection.

While our method applies to the i.i.d. situation, we provided a hint of what can be expected in a non-i.i.d. analysis of randomness from post-selected data. It would be interesting to know whether the bounds presented here remain true in this case, if no information about the outputs is revealed to the adversary.

More generally, our method could be used to analyse the randomness present in disjoint subsets of events independently. As shown in 4.3.4(b), this may lead to tighter randomness estimates than previously achievable, solving in particular the two-day paradox. We leave the full generalization of this result for further study.

CHAPTER 5

CONCLUSIONS AND OUTLOOK

In this thesis we have presented several results in the field of quantum key distribution and quantum randomness generation. The main workforce behind the security of these applications is the unique properties of quantum mechanical systems: measurement disturbance, no cloning, nonlocality.

We presented a possible extension of the reference-frame-independent protocol to d-level quantum signals and showed that the protocol is actually tomographic in nature; it is better to use directly the different correlations, rather than compressing them into a single frame independent parameter. We also proposed a framework to prove the security of distributed-phase-reference protocols which is based on the de Finetti theorem. This is only a first step towards a full framework tailored specifically for the DPR protocols, which is a fruitful direction to explore.

We studied the relationship between amount of extractable randomness and the levels of characterization of the devices in the setting of measurement independence. Since trust or levels of characterization is a personal issue, one cannot force the users into a certain level of trust. In fact, they must choose a level of trust on the devices with which they are comfortable, and then our results give the corresponding amount of random bits which can be extracted by two universal hashing. Our frame work here is applicable to any scenario where guessing probability and trust is involved, e.g. two party cryptographic applications.

Leaving measurement independent behind, we explained the important role of the randomness of the inputs in Bell tests: they are required to reveal the local behavior of the boxes (if any) in an adversarial scenario. This has important consequences for randomness amplification, namely that Bell tests cannot be used to amplify arbitrary min-entropy source in the device-independent scenario. Moreover, the main concern with any randomness amplification protocol is to provide a

reasonable estimate for the initial quality of the source (which is relative to Eve and is therefore unobservable by definition). This is the most pressing question that randomness amplification commynity must answer in order for the field to be meaningful. Perhaps some further fundamental physical principles not present in quantum theory can provide us with such a bound.

The last topic we discussed in this thesis is the amount of randomness in a post-selected data of Bell tests. It is well known that improper treatment of the experimental data opens up the detection loophole (i.e. an adversary can fake a violation of Bell inequalities with local variables); we presented a way to analyze the data for the desired amount of randomness without opening such loophole. The result is applicable to an experiment, especially in photonic implementations with source and detector inefficiencies.

Our randomness results are not presented in the most general framework of non-i.i.d adversaries. Unless there is any physical reason limiting the adversary's strategy (such as the de Finetti theorem in QKD), we should aim any cryptographic security results toward the most powerful model allowed by the current physical theory. In fact, beyond i.i.d. in quantum information is currently a hot topic of research [143]. Therefore, an extension of our results to the non-i.i.d. adversary is an interesting direction to explore. Moreover, our last work on the randomness in post-selected data as well as earlier works on QKD [20] have also hinted the possibility of pre-processing as a way to enhance the performance of quantum information processing tasks.

BIBLIOGRAPHY

[1] G. A. D. Briggs, J. N. Butterfield, and A. Zeilinger. "The Oxford Questions on the foundations of quantum physics". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 469.2157 (2013). ISSN: 1364-5021. DOI: 10.1098/rspa.2013.0299.

[2] G. Brassard. "Brief history of quantum cryptography: a personal perspective". In: *Theory and Practice in Information-Theoretic Security, 2005. IEEE Information Theory Workshop on*. Oct. 2005, pp. 19–23. DOI: 10.1109/ITWTPI.2005.1543949.

[3] Charles H. Bennett and Gilles Brassard. In: *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*. New York: IEEE, 1984, pp. 175–179.

[4] A. Einstein, B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Phys. Rev.* 47 (10 May 1935), pp. 777–780. DOI: 10.1103/PhysRev.47.777. URL: http://link.aps.org/doi/10.1103/PhysRev.47.777.

[5] John S. Bell. "On the Einstein Podolsky Rosen Paradox". In: *Physics (Long Island City, N.Y.)* 1.3 (Nov. 1964), p. 195.

[6] Stuart J. Freedman and John F. Clauser. "Experimental Test of Local Hidden-Variable Theories". In: *Phys. Rev. Lett.* 28 (14 Apr. 1972), pp. 938–941. DOI: 10.1103/PhysRevLett.28.938. URL: http://link.aps.org/doi/10.1103/PhysRevLett.28.938.

[7] Alain Aspect, Philippe Grangier, and Gérard Roger. "Experimental Realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment* : A New Violation of Bell's Inequalities". In: *Phys. Rev. Lett.* 49 (2 July 1982), pp. 91–

94. DOI: 10.1103/PhysRevLett.49.91. URL: http://link.aps.org/doi/10.1103/PhysRevLett.49.91.

[8] Nicolas Brunner et al. "Bell nonlocality". In: *Reviews of Modern Physics* 86.419 (2014), p. 012311.

[9] Simon Singh. *The Code Book: The Evolution of Secrecy from Mary, Queen of Scots, to Quantum Cryptography*. 1st. New York, NY, USA: Doubleday, 1999. ISBN: 0385495315.

[10] R. L. Rivest, A. Shamir, and L. Adleman. "A Method for Obtaining Digital Signatures and Public-key Cryptosystems". In: *Commun. ACM* 21.2 (Feb. 1978), pp. 120–126. ISSN: 0001-0782. DOI: 10.1145/359340.359342. URL: http://doi.acm.org/10.1145/359340.359342.

[11] T. D. Ladd et al. "Quantum computers". In: *Nature* 464.7285 (Mar. 2010), pp. 45–53. ISSN: 0028-0836. DOI: 10.1038/nature08812. URL: http://dx.doi.org/10.1038/nature08812.

[12] C.E. Shannon. "Communication theory of secrecy systems". In: *Bell System Technical Journal, The* 28.4 (Oct. 1949), pp. 656–715. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1949.tb00928.x.

[13] R. Colbeck. "Quantum And Relativistic Protocols For Secure Multi-Party Computation". PhD thesis. PhD Thesis, 2009, Nov. 2009.

[14] Dagmar BruSS. "Optimal Eavesdropping in Quantum Cryptography with Six States". In: *Phys. Rev. Lett.* 81 (14 Oct. 1998), pp. 3018–3021. DOI: 10.1103/PhysRevLett.81.3018. URL: http://link.aps.org/doi/10.1103/PhysRevLett.81.3018.

[15] H. Bechmann-Pasquinucci and N. Gisin. "Incoherent and coherent eavesdropping in the six-state protocol of quantum cryptography". In: *Phys. Rev. A* 59 (6 June 1999), pp. 4238–4248. DOI: 10.1103/PhysRevA.59.4238. URL: http://link.aps.org/doi/10.1103/PhysRevA.59.4238.

[16] Dominic Mayers. "Quantum Key Distribution and String Oblivious Transfer in Noisy Channels". In: *Proceedings of the 16th Annual International Cryptology Conference on Advances in Cryptology*. CRYPTO '96. London, UK, UK: Springer-Verlag, 1996, pp. 343–357. ISBN: 3-540-61512-1. URL: http://dl.acm.org/citation.cfm?id=646761.706026.

[17] Dominic Mayers. "Unconditional Security in Quantum Cryptography". In: *J. ACM* 48.3 (May 2001), pp. 351–406. ISSN: 0004-5411. DOI: 10.1145/382780.382781. URL: http://doi.acm.org/10.1145/382780.382781.

[18] Hoi-Kwong Lo and H. F. Chau. "Unconditional Security of Quantum Key Distribution over Arbitrarily Long Distances". In: *Science* 283.5410 (1999), pp. 2050–2056. DOI: 10.1126/science.283.5410.2050.

[19] Peter W. Shor and John Preskill. "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol". In: *Phys. Rev. Lett.* 85 (2 July 2000), pp. 441–444. DOI: 10.1103/PhysRevLett.85.441. URL: http://link.aps.org/doi/10.1103/PhysRevLett.85.441.

[20] B. Kraus, N. Gisin, and R. Renner. "Lower and Upper Bounds on the Secret-Key Rate for Quantum Key Distribution Protocols Using One-Way Classical Communication". In: *Phys. Rev. Lett.* 95 (8 Aug. 2005), p. 080501. DOI: 10.1103/PhysRevLett.95.080501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.95.080501.

[21] Renato Renner, Nicolas Gisin, and Barbara Kraus. "Information-theoretic security proof for quantum-key-distribution protocols". In: *Phys. Rev. A* 72 (1 July 2005), p. 012332. DOI: 10.1103/PhysRevA.72.012332. URL: http://link.aps.org/doi/10.1103/PhysRevA.72.012332.

[22] Renato Renner and Robert König. "Universally Composable Privacy Amplification Against Quantum Adversaries". English. In: *Theory of Cryptography*. Ed. by Joe Kilian. Vol. 3378. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, pp. 407–425. ISBN: 978-3-540-24573-5. DOI: 10.1007/978-3-540-30576-7_22. URL: http://dx.doi.org/10.1007/978-3-540-30576-7_22.

[23] Renato Renner. "Security of quantum key distribution". In: *International Journal of Quantum Information* 06.01 (2008), pp. 1–127. DOI: 10.1142/S0219749908003256. eprint: http://www.worldscientific.com/doi/pdf/10.1142/S0219749908003256. URL: http://www.worldscientific.com/doi/abs/10.1142/S0219749908003256.

[24] Matthias Christandl, Robert König, and Renato Renner. "Postselection Technique for Quantum Channels with Applications to Quantum Cryptography". In: *Phys. Rev. Lett.* 102 (2 Jan. 2009), p. 020504. DOI: 10.1103/PhysRevLett.102.020504. URL: http://link.aps.org/doi/10.1103/PhysRevLett.102.020504.

[25] Marco Tomamichel et al. "Tight finite-key analysis for quantum cryptography". In: *Nat Commun* 3 (Jan. 2012), p. 634. DOI: 10.1038/ncomms1631. URL: http://dx.doi.org/10.1038/ncomms1631.

[26] Hoi-Kwong Lo, Marcos Curty, and Bing Qi. "Measurement-Device-Independent Quantum Key Distribution". In: *Physical Review Letters* 108 (2012), p. 130503.

[27] Antonio Acín et al. "Device-Independent Security of Quantum Cryptography against Collective Attacks". In: *Phys. Rev. Lett.* 98 (23 June 2007), p. 230501. DOI: 10.1103/PhysRevLett.98.230501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.98.230501.

[28] Stefano Pironio et al. "Random numbers certified by Bell's theorem". In: *Nature* 464.7291 (2010), pp. 1021–1024.

[29] U. V. Vazirani and T. Vidick. "Certifiable Quantum Dice - Or, testable exponential randomness expansion". In: *ArXiv e-prints* (Nov. 2011). arXiv: 1111.6054 [quant-ph].

[30] M. Coudron and H. Yuen. "Infinite Randomness Expansion and Amplification with a Constant Number of Devices". In: *ArXiv e-prints* (Oct. 2013). arXiv: 1310.6755 [quant-ph].

[31] C. A. Miller and Y. Shi. "Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices". In: *ArXiv e-prints* (Feb. 2014). arXiv: 1402.0489 [quant-ph].

[32] Roger Colbeck and Renato Renner. "Free randomness can be amplified". In: *Nature Physics* 8.6 (2012), pp. 450–453.

[33] Rodrigo Gallego et al. "Full randomness from arbitrarily deterministic events". In: *Nat Commun* 4 (Oct. 2013). URL: http://dx.doi.org/10.1038/ncomms3654.

[34] Matej Pivoluska and Martin Plesch. "Device Independent Random Number Generation". In: *Acta Physica Slovaca* 64 (6 Dec. 2014), pp. 601–664. DOI: 10.2478/apsrt-2014-0006. URL: http://link.aps.org/doi/10.1103/PhysRevLett.85.441.

[35] A. Laing et al. In: *Phys. Rev. A* 82 (2010), p. 012304.

[36] Edo Waks, Hiroki Takesue, and Yoshihisa Yamamoto. "Security of differential-phase-shift quantum key distribution against individual attacks". In: *Phys. Rev. A* 73 (1 Jan. 2006), p. 012344. DOI: 10.1103/PhysRevA.73.012344. URL: http://link.aps.org/doi/10.1103/PhysRevA.73.012344.

[37] Cyril Branciard, Nicolas Gisin, and Valerio Scarani. "Upper bounds for the security of two distributed-phase reference protocols of quantum cryptography". In: *New Journal of Physics* 10.1 (2008), p. 013031. URL: http://stacks.iop.org/1367-2630/10/i=1/a=013031.

[38] Kai Wen, Kiyoshi Tamaki, and Yoshihisa Yamamoto. "Unconditional Security of Single-Photon Differential Phase Shift Quantum Key Distribution". In: *Phys. Rev. Lett.* 103 (17 Oct. 2009), p. 170503. DOI: 10.1103/PhysRevLett.103.170503. URL: http://link.aps.org/doi/10.1103/PhysRevLett.103.170503.

[39] J. Bouda et al. "Device-independent randomness extraction for arbitrarily weak min-entropy source". In: *ArXiv e-prints* (Feb. 2014). arXiv: 1402.0974 [quant-ph].

[40] Le Phuc Thinh, Lana Sheridan, and Valerio Scarani. "Tomographic quantum cryptography protocols are reference frame independent". In: *International Journal of Quantum Information* 10.03 (2012), p. 1250035. DOI: 10.1142/S0219749912500359. eprint: http://www.worldscientific.com/doi/pdf/10.1142/S0219749912500359. URL: http://www.worldscientific.com/doi/abs/10.1142/S0219749912500359.

[41] Tobias Moroder et al. "Security of Distributed-Phase-Reference Quantum Key Distribution". In: *Phys. Rev. Lett.* 109 (26 Dec. 2012), p. 260501. DOI: 10.1103/PhysRevLett.109.260501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.109.260501.

[42] Yun Zhi Law et al. "Quantum randomness extraction for various levels of characterization of the devices". In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (2014), p. 424028. URL: http://stacks.iop.org/1751-8121/47/i=42/a=424028.

[43] Le Phuc Thinh, Lana Sheridan, and Valerio Scarani. "Bell tests with min-entropy sources". In: *Phys. Rev. A* 87 (6 June 2013), p. 062121. DOI: 10.1103/PhysRevA.87.062121. URL: http://link.aps.org/doi/10.1103/PhysRevA.87.062121.

[44] L. Phuc Thinh et al. "Randomness in post-selected events". In: *ArXiv e-prints* (June 2015). arXiv: 1506.03953 [quant-ph].

[45] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information*. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, 2000. ISBN: 9780521635035. URL: http://books.google.com.sg/books?id=65FqEKQOfP8C.

[46] Marco Tomamichel. "A framework for non-asymptotic quantum information theory". In: *arXiv preprint arXiv:1203.2142* (2012).

[47] R. Konig, R. Renner, and C. Schaffner. "The Operational Meaning of Min- and Max-Entropy". In: *Information Theory, IEEE Transactions on* 55.9 (Sept. 2009), pp. 4337–4347. ISSN: 0018-9448. DOI: 10.1109/TIT.2009.2025545.

[48] Miguel Navascués, Stefano Pironio, and Antonio Acín. "Bounding the Set of Quantum Correlations". In: *Phys. Rev. Lett.* 98 (1 Jan. 2007), p. 010401. DOI: 10.1103/PhysRevLett.98.010401. URL: http://link.aps.org/doi/10.1103/PhysRevLett.98.010401.

[49] Miguel Navascués, Stefano Pironio, and Antonio Acín. "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations". In: *New Journal of Physics* 10.7 (2008), p. 073013. URL: http://stacks.iop.org/1367-2630/10/i=7/a=073013.

[50] M.S. Sharbaf. "Quantum cryptography: An emerging technology in network security". In: *Technologies for Homeland Security (HST), 2011 IEEE International Conference on*. Nov. 2011, pp. 13–19. DOI: 10.1109/THS.2011.6107841.

[51] C. Portmann and R. Renner. "Cryptographic security of quantum key distribution". In: *ArXiv e-prints* (Sept. 2014). arXiv: 1409.3525 [quant-ph].

[52] Stephen D. Bartlett, Terry Rudolph, and Robert W. Spekkens. "Reference frames, superselection rules, and quantum information". In: *Rev. Mod. Phys.* 79 (2 Apr. 2007), pp. 555–609. DOI: 10.1103/RevModPhys.79.555. URL: http://link.aps.org/doi/10.1103/RevModPhys.79.555.

[53] J.-C. Boileau et al. "Robust Polarization-Based Quantum Key Distribution over a Collective-Noise Channel". In: *Phys. Rev. Lett.* 92 (1 Jan. 2004), p. 017901. DOI: 10.1103/PhysRevLett.92.017901. URL: http://link.aps.org/doi/10.1103/PhysRevLett.92.017901.

[54] L. Aolita and S. P. Walborn. "Quantum Communication without Alignment using Multiple-Qubit Single-Photon States". In: *Phys. Rev. Lett.* 98 (10 Mar. 2007), p. 100501. DOI: 10.1103/PhysRevLett.98.100501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.98.100501.

[55] Renato Renner. "Symmetry of large physical systems implies independence of subsystems". In: *Nat Phys* 3.9 (Sept. 2007), pp. 645–649. ISSN: 1745-2473. DOI: 10.1038/nphys684. URL: http://dx.doi.org/10.1038/nphys684.

[56] Yeong Cherng Liang et al. "Tomographic quantum cryptography". In: *Phys. Rev. A* 68 (2 Aug. 2003), p. 022324. DOI: 10.1103/PhysRevA.68.022324. URL: http://link.aps.org/doi/10.1103/PhysRevA.68.022324.

[57] Kyo Inoue, Edo Waks, and Yoshihisa Yamamoto. "Differential Phase Shift Quantum Key Distribution". In: *Phys. Rev. Lett.* 89 (3 June 2002), p. 037902. DOI: 10.1103/PhysRevLett.89.037902. URL: http://link.aps.org/doi/10.1103/PhysRevLett.89.037902.

[58] Nicolas Gisin et al. "Towards practical and fast quantum cryptography". In: *arXiv preprint quant-ph/0411022* (2004).

[59] Normand J. Beaudry, Tobias Moroder, and Norbert Lütkenhaus. "Squashing Models for Optical Measurements in Quantum Communication". In: *Phys. Rev. Lett.* 101 (9 Aug. 2008), p. 093601. DOI: 10.1103/PhysRevLett.101.093601. URL: http://link.aps.org/doi/10.1103/PhysRevLett.101.093601.

[60] Toyohiro Tsurumaru and Kiyoshi Tamaki. "Security proof for quantum-key-distribution systems with threshold detectors". In: *Phys. Rev. A* 78 (3 Sept. 2008), p. 032302. DOI: 10.1103/PhysRevA.78.032302. URL: http://link.aps.org/doi/10.1103/PhysRevA.78.032302.

[61] Daniel Gottesman et al. "Security of quantum key distribution with imperfect devices". In: *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*. IEEE. 2004, p. 136.

[62] Hoi-Kwong Lo and John Preskill. "Security of Quantum Key Distribution Using Weak Coherent States with Nonrandom Phases". In: *Quantum Info. Comput.* 7.5 (July 2007), pp. 431–458. ISSN: 1533-7146. URL: http://dl.acm.org/citation.cfm?id=2011832.2011834.

[63] Valerio Scarani et al. "Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations". In: *Phys. Rev. Lett.* 92 (5 Feb. 2004), p. 057901. DOI: 10.1103/PhysRevLett.92.057901. URL: http://link.aps.org/doi/10.1103/PhysRevLett.92.057901.

[64] Cyril Branciard et al. "Zero-Error Attacks and Detection Statistics in the Coherent One-Way Protocol for Quantum Cryptography". In: *Quantum Information and Computation* 7.7 (2007), pp. 639–664.

[65] A. Grudka et al. "Free randomness amplification using bipartite chain correlations". In: *ArXiv e-prints* (Mar. 2013). arXiv: 1303.5591 [quant-ph].

[66] R. Ramanathan et al. "Robust Device Independent Randomness Amplification". In: *ArXiv e-prints* (Aug. 2013). arXiv: 1308.4635 [quant-ph].

[67] F. G. S. L. Brandao et al. "Robust Device-Independent Randomness Amplification with Few Devices". In: *ArXiv e-prints* (Oct. 2013). arXiv: 1310.4544 [quant-ph].

[68] Piotr Mironowicz, Rodrigo Gallego, and Marcin Pawowski. "Robust amplification of Santha-Vazirani sources with three devices". In: *Phys. Rev. A* 91 (3 Mar. 2015), p. 032317. DOI: 10.1103/PhysRevA.91.032317. URL: http://link.aps.org/doi/10.1103/PhysRevA.91.032317.

[69] K.-M. Chung, Y. Shi, and X. Wu. "Physical Randomness Extractors: Generating Random Numbers with Minimal Assumptions". In: *ArXiv e-prints* (Feb. 2014). arXiv: 1402.4797 [quant-ph].

[70] Antonio Ac´n, Nicolas Gisin, and Lluis Masanes. "From Bell's Theorem to Secure Quantum Key Distribution". In: *Phys. Rev. Lett.* 97 (12 Sept. 2006), p. 120405. DOI: 10.1103/PhysRevLett.97.120405. URL: http://link.aps.org/doi/10.1103/PhysRevLett.97.120405.

[71] Ivan B Damgård et al. "Cryptography in the bounded-quantum-storage model". In: *SIAM Journal on Computing* 37.6 (2008), pp. 1865–1890.

[72] Christian Schaffner, Barbara Terhal, and Stephanie Wehner. "Robust Cryptography in the Noisy-quantum-storage Model". In: *Quantum Info. Comput.* 9.11 (Nov. 2009), pp. 963–996. ISSN: 1533-7146. URL: http://dl.acm.org/citation.cfm?id=2012098.2012102.

[73] M. Coudron, T. Vidick, and H. Yuen. "Robust Randomness Amplifiers: Upper and Lower Bounds". In: *ArXiv e-prints* (2013). arXiv: 1305.6626 [quant-ph].

[74] Jonathan Barrett, Roger Colbeck, and Adrian Kent. "Memory Attacks on Device-Independent Quantum Cryptography". In: *Phys. Rev. Lett.* 110 (1 Jan. 2013), p. 010503. DOI: 10.1103/PhysRevLett.110.010503. URL: http://link.aps.org/doi/10.1103/PhysRevLett.110.010503.

[75] Stefano Pironio and Serge Massar. "Security of practical private randomness generation". In: *Phys. Rev. A* 87 (1 Jan. 2013), p. 012336. DOI: 10.1103/PhysRevA.87.012336. URL: http://link.aps.org/doi/10.1103/PhysRevA.87.012336.

[76] Serge Fehr, Ran Gelles, and Christian Schaffner. "Security and composability of randomness expansion from Bell inequalities". In: *Phys. Rev. A* 87 (1 Jan. 2013), p. 012335. DOI: 10.1103/PhysRevA.87.012335. URL: http://link.aps.org/doi/10.1103/PhysRevA.87.012335.

[77] Daniel F. V. James et al. "Measurement of qubits". In: *Phys. Rev. A* 64 (5 Oct. 2001), p. 052312. DOI: 10.1103/PhysRevA.64.052312. URL: http://link.aps.org/doi/10.1103/PhysRevA.64.052312.

[78] Tobias Moroder et al. "Entanglement verification with realistic measurement devices via squashing operations". In: *Phys. Rev. A* 81 (5 May 2010), p. 052342. DOI: 10.1103/PhysRevA.81.052342. URL: http://link.aps.org/doi/10.1103/PhysRevA.81.052342.

[79] Yong Siah Teo et al. "Incomplete quantum state estimation: A comprehensive study". In: *Phys. Rev. A* 85 (4 Apr. 2012), p. 042317. DOI: 10.1103/PhysRevA.85.042317. URL: http://link.aps.org/doi/10.1103/PhysRevA.85.042317.

[80] Denis Rosset et al. "Imperfect measurement settings: Implications for quantum state tomography and entanglement witnesses". In: *Phys. Rev. A* 86 (6 Dec. 2012), p. 062325. DOI: 10.1103/PhysRevA.86.062325. URL: http://link.aps.org/doi/10.1103/PhysRevA.86.062325.

[81] Lars Lydersen et al. "Hacking commercial quantum cryptography systems by tailored bright illumination". In: *Nat Photon* 4.10 (Oct. 2010), pp. 686–689. ISSN: 1749-4885. DOI: 10.1038/nphoton.2010.214. URL: http://dx.doi.org/10.1038/nphoton.2010.214.

[82] Qin Liu et al. "A universal setup for active control of a single-photon detector". In: *Review of Scientific Instruments* 85.1, 013108 (2014). DOI: http://dx.doi.org/10.1063/1.4854615. URL: http://scitation.aip.org/content/aip/journal/rsi/85/1/10.1063/1.4854615.

[83] Michael G. Tanner, Vadim Makarov, and Robert H. Hadfield. "Optimised quantum hacking of superconducting nanowire single-photon detectors". In: *Opt. Express* 22.6 (Mar. 2014), pp. 6734–6748. DOI: 10.1364/OE.22.006734. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-22-6-6734.

[84] Michael J. W. Hall. "Relaxed Bell inequalities and Kochen-Specker theorems". In: *Phys. Rev. A* 84 (2 Aug. 2011), p. 022102. DOI: 10.1103/PhysRevA.84.022102. URL: http://link.aps.org/doi/10.1103/PhysRevA.84.022102.

[85] Marcin Pawlowski and Nicolas Brunner. "Semi-device-independent security of one-way quantum key distribution". In: *Physical Review A* 84.1 (2011), 010203(R).

[86] Cyril Branciard et al. "One-sided device-independent quantum key distribution: Security, feasibility, and the connection with steering". In: *Phys. Rev. A* 85 (1 Jan. 2012), p. 010301. DOI: 10.1103/PhysRevA.85.010301. URL: http://link.aps.org/doi/10.1103/PhysRevA.85.010301.

[87] Jean-Daniel Bancal, Lana Sheridan, and Valerio Scarani. "More randomness from the same data". In: *New Journal of Physics* 16.3 (2014), p. 033011. URL: http://stacks.iop.org/1367-2630/16/i=3/a=033011.

[88] Marissa Giustina et al. "Bell violation using entangled photons without the fair-sampling assumption". In: *Nature* 497.7448 (May 2013), pp. 227–230. ISSN: 0028-0836. URL: http://dx.doi.org/10.1038/nature12012.

[89] B. G. Christensen et al. "Detection-Loophole-Free Test of Quantum Nonlocality, and Applications". In: *Phys. Rev. Lett.* 111 (13 Sept. 2013), p. 130406. DOI: 10.1103/PhysRevLett.111.130406. URL: http://link.aps.org/doi/10.1103/PhysRevLett.111.130406.

[90] R. D. Gill. "Time, Finite Statistics, and Bell's Fifth Position". In: *eprint arXiv:quant-ph/0301059* (Jan. 2003). eprint: quant-ph/0301059.

[91] Yanbao Zhang, Scott Glancy, and Emanuel Knill. "Asymptotically optimal data analysis for rejecting local realism". In: *Phys. Rev. A* 84 (6 Dec. 2011), p. 062118. DOI: 10.1103/PhysRevA.84.062118. URL: http://link.aps.org/doi/10.1103/PhysRevA.84.062118.

[92] Antonio Acín, Serge Massar, and Stefano Pironio. "Randomness versus Nonlocality and Entanglement". In: *Phys. Rev. Lett.* 108 (10 Mar. 2012), p. 100402. DOI: 10.1103/PhysRevLett.108.100402. URL: http://link.aps.org/doi/10.1103/PhysRevLett.108.100402.

[93] O Nieto-Silleras, S Pironio, and J Silman. "Using complete measurement statistics for optimal device-independent randomness evaluation". In: *New Journal of Physics* 16.1 (2014), p. 013035. URL: http://stacks.iop.org/1367-2630/16/i=1/a=013035.

[94] Marco Tomamichel and Renato Renner. "Uncertainty Relation for Smooth Entropies". In: *Phys. Rev. Lett.* 106 (11 Mar. 2011), p. 110506. DOI: 10.1103/PhysRevLett.106.110506. URL: http://link.aps.org/doi/10.1103/PhysRevLett.106.110506.

[95] Steve James Jones, Howard Mark Wiseman, and Andrew C Doherty. "Entanglement, Einstein-Podolsky-Rosen correlations, Bell nonlocality, and steering". In: *Physical Review A* 76.5 (2007), p. 052116.

[96] Eric Gama Cavalcanti et al. "Experimental criteria for steering and the Einstein-Podolsky-Rosen paradox". In: *Physical Review A* 80.3 (2009), p. 032112.

[97] Sandu Popescu and Daniel Rohrlich. "Which states violate Bell's inequality maximally?" In: *Physics Letters A* 169.6 (1992), pp. 411–414.

[98] Dylan John Saunders et al. "Experimental EPR-steering using Bell-local states". In: *Nature Physics* 6.11 (2010), pp. 845–849.

[99] Hans Maassen and J.B.M. Uffink. "Generalized entropic uncertainty relations". In: *Phys. Rev. Lett.* 60.12 (1988), p. 1103.

[100] Ping-Xing Chen et al. "Ancilla dimensions needed to carry out positive-operator-valued measurement". In: *Phys. Rev. A* 76 (6 Dec. 2007), p. 060303. DOI: 10.1103/PhysRevA.76.060303. URL: http://link.aps.org/doi/10.1103/PhysRevA.76.060303.

[101] M. Fiorentino et al. "Secure self-calibrating quantum random-bit generator". In: *Phys. Rev. A* 75 (3 Mar. 2007), p. 032334. DOI: 10.1103/PhysRevA.75.032334. URL: http://link.aps.org/doi/10.1103/PhysRevA.75.032334.

[102] G. Vallone et al. "Self-calibrating quantum random number generator based on the uncertainty principle". In: *arXiv preprint arXiv:1401.7917* (2014).

[103] Thomas Jennewein et al. "A fast and compact quantum random number generator". In: *Review of Scientific Instruments* 71.4 (2000), pp. 1675–1680. DOI: http://dx.doi.org/10.1063/1.1150518. URL: http://scitation.aip.org/content/aip/journal/rsi/71/4/10.1063/1.1150518.

[104] André Stefanov et al. "Optical quantum random number generator". In: *Journal of Modern Optics* 47.4 (2000), pp. 595–598. DOI: 10.1080/09500340008233380. eprint: http://dx.doi.org/10.1080/09500340008233380. URL: http://dx.doi.org/10.1080/09500340008233380.

[105] T. Symul, S.M. Assad, and P.K. Lam. "Real time demonstration of high bitrate quantum random number generation with coherent laser light". In: *Appl. Phys. Lett.* 98 (2011), p. 231103.

[106] Dominic Mayers and Andrew Yao. "Quantum Cryptography with Imperfect Apparatus". In: *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*. FOCS '98. Washington, DC, USA: IEEE Computer Society, 1998, pp. 503–. ISBN: 0-8186-9172-7. URL: http://dl.acm.org/citation.cfm?id=795664.796390.

[107] Umesh Vazirani and Thomas Vidick. "Fully Device-Independent Quantum Key Distribution". In: *Phys. Rev. Lett.* 113 (14 Sept. 2014), p. 140501. DOI: 10.1103/PhysRevLett.113.140501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.113.140501.

[108] C.-E. Bardyn et al. "Device-independent state estimation based on Bell's inequalities". In: *Phys. Rev. A* 80 (6 Dec. 2009), p. 062327. DOI: 10.1103/PhysRevA.80.062327. URL: http://link.aps.org/doi/10.1103/PhysRevA.80.062327.

[109] M McKague, T H Yang, and V Scarani. "Robust self-testing of the singlet". In: *Journal of Physics A: Mathematical and Theoretical* 45.45 (2012), p. 455304. URL: http://stacks.iop.org/1751-8121/45/i=45/a=455304.

[110] Rafael Rabelo et al. "Device-Independent Certification of Entangled Measurements". In: *Phys. Rev. Lett.* 107 (5 July 2011), p. 050502. DOI: 10.1103/PhysRevLett.107.050502. URL: http://link.aps.org/doi/10.1103/PhysRevLett.107.050502.

[111] S. Pironio et al. In: *Nature* 464 (2010), p. 1021.

[112] Roger Colbeck and Adrian Kent. "Private randomness expansion with untrusted devices". In: *Journal of Physics A: Mathematical and Theoretical* 44.9 (2011), p. 095305.

[113] Dax Enshan Koh et al. "Effects of Reduced Measurement Independence on Bell-Based Randomness Expansion". In: *Phys. Rev. Lett.* 109 (16 Oct. 2012), p. 160404. DOI: 10.1103/PhysRevLett.109.160404. URL: http://link.aps.org/doi/10.1103/PhysRevLett.109.160404.

[114] Piotr Mironowicz and Marcin Pawowski. "Robustness of quantum-randomness expansion protocols in the presence of noise". In: *Phys. Rev. A* 88 (3 Sept. 2013), p. 032319. DOI: 10.1103/PhysRevA.88.032319. URL: http://link.aps.org/doi/10.1103/PhysRevA.88.032319.

[115] Jonathan Barrett et al. "Quantum nonlocality, Bell inequalities, and the memory loophole". In: *Phys. Rev. A* 66 (4 Oct. 2002), p. 042111. DOI: 10.1103/PhysRevA.66.042111. URL: http://link.aps.org/doi/10.1103/PhysRevA.66.042111.

[116] Jonathan Barrett, Adrian Kent, and Stefano Pironio. "Maximally Nonlocal and Monogamous Quantum Correlations". In: *Phys. Rev. Lett.* 97 (17 Oct. 2006), p. 170409. DOI: 10.1103/PhysRevLett.97.170409. URL: http://link.aps.org/doi/10.1103/PhysRevLett.97.170409.

[117] Roger Colbeck and Renato Renner. "Hidden Variable Models for Quantum Theory Cannot Have Any Local Part". In: *Phys. Rev. Lett.* 101 (5 Aug. 2008), p. 050403. DOI: 10.1103/PhysRevLett.101.050403. URL: http://link.aps.org/doi/10.1103/PhysRevLett.101.050403.

[118] Salil P. Vadhan. "Pseudorandomness". In: *Foundations and Trendső in Theoretical Computer Science* 7.13 (2011), pp. 1–336. ISSN: 1551-305X. DOI: 10.1561/0400000010. URL: http://dx.doi.org/10.1561/0400000010.

[119] Miklos Santha and Umesh V. Vazirani. "Generating quasi-random sequences from semi-random sources". In: *Journal of Computer and System Sciences* 33.1 (1986), pp. 75–87. ISSN: 0022-0000. DOI: http://dx.doi.org/10.1016/0022-0000(86)90044-9. URL: http://www.sciencedirect.com/science/article/pii/0022000086900449.

[120] M. Pawlowski et al. "When non i.i.d. information sources can be communicationally useful?" In: *ArXiv e-prints* (Feb. 2009). arXiv: 0902.2162 [quant-ph].

[121] Arthur Fine. "Hidden Variables, Joint Probability, and the Bell Inequalities". In: *Phys. Rev. Lett.* 48 (5 Feb. 1982), pp. 291–295. DOI: 10.1103/PhysRevLett.48.291. URL: http://link.aps.org/doi/10.1103/PhysRevLett.48.291.

[122] Stefan Zohren and Richard D. Gill. "Maximal Violation of the Collins-Gisin-Linden-Massar-Popescu Inequality for Infinite Dimensional States". In: *Phys. Rev. Lett.* 100 (12 Mar. 2008), p. 120406. DOI: 10.1103/PhysRevLett.100.120406. URL: http://link.aps.org/doi/10.1103/PhysRevLett.100.120406.

[123] S. Zohren et al. "A tight Tsirelson inequality for infinitely many outcomes". In: *EPL (Europhysics Letters)* 90.1 (2010), p. 10002. URL: http://stacks.iop.org/0295-5075/90/i=1/a=10002.

[124] N. David Mermin. "Extreme quantum entanglement in a superposition of macroscopically distinct states". In: *Phys. Rev. Lett.* 65 (15 Oct. 1990), pp. 1838–1840. DOI: 10.1103/PhysRevLett.65.1838. URL: http://link.aps.org/doi/10.1103/PhysRevLett.65.1838.

[125] A V Belinskii and D N Klyshko. "Interference of light and Bell's theorem". In: *Physics-Uspekhi* 36.8 (1993), p. 653. URL: http://stacks.iop.org/1063-7869/36/i=8/a=R01.

[126] Adán Cabello, Otfried Gühne, and David Rodríguez. "Mermin inequalities for perfect correlations". In: *Phys. Rev. A* 77 (6 June 2008), p. 062106. DOI: 10.1103/PhysRevA.77.062106. URL: http://link.aps.org/doi/10.1103/PhysRevA.77.062106.

[127] Gilles Pütz et al. "Arbitrarily Small Amount of Measurement Independence Is Sufficient to Manifest Quantum Nonlocality". In: *Phys. Rev. Lett.* 113 (19 Nov. 2014), p. 190402. DOI: 10.1103/PhysRevLett.113.190402. URL: http://link.aps.org/doi/10.1103/PhysRevLett.113.190402.

[128] R. Impagliazzo, L. Levin, and M. Luby. In: *STOC '89 Proceedings of the twenty-first annual ACM symposium on Theory of computing*. 1989, pp. 12–24.

[129] Y. Dodis and J. Spencer. In: *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*. 2002, pp. 376–385.

[130] Y. Dodis et al. In: *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*. 2004, pp. 196 –205.

[131] W. Tittel et al. "Violation of Bell Inequalities by Photons More Than 10 km Apart". In: *Phys. Rev. Lett.* 81 (17 Oct. 1998), pp. 3563–3566. DOI: 10.1103/PhysRevLett.81.3563. URL: http://link.aps.org/doi/10.1103/PhysRevLett.81.3563.

[132] Gregor Weihs et al. "Violation of Bell's Inequality under Strict Einstein Locality Conditions". In: *Phys. Rev. Lett.* 81 (23 Dec. 1998), pp. 5039–5043. DOI: 10.1103/PhysRevLett.81.5039. URL: http://link.aps.org/doi/10.1103/PhysRevLett.81.5039.

[133] Philippe H. Eberhard. "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment". In: *Phys. Rev. A* 47 (2 Feb. 1993), R747–R750. DOI: 10.1103/PhysRevA.47.R747. URL: http://link.aps.org/doi/10.1103/PhysRevA.47.R747.

[134] N. David Mermin. "The EPR ExperimentThoughts about the Loophole". In: *Annals of the New York Academy of Sciences* 480.1 (1986), pp. 422–427. ISSN: 1749-6632. DOI: 10.1111/j.1749-6632.1986.tb12444.x. URL: http://dx.doi.org/10.1111/j.1749-6632.1986.tb12444.x.

[135] Stefano Pironio and Serge Massar. "Security of practical private randomness generation". In: *Phys. Rev. A* 87 (1 Jan. 2013), p. 012336. DOI: 10.1103/PhysRevA.87.012336. URL: http://link.aps.org/doi/10.1103/PhysRevA.87.012336.

[136] S. Pironio, M. Navascués, and A. Acín. "Convergent Relaxations of Polynomial Optimization Problems with Noncommuting Variables". In: *SIAM Journal on Optimization* 20.5 (2010), pp. 2157–2180. DOI: 10.1137/090760155. eprint: http://dx.doi.org/10.1137/090760155. URL: http://dx.doi.org/10.1137/090760155.

[137] Tobias Moroder et al. "Device-Independent Entanglement Quantification and Related Applications". In: *Phys. Rev. Lett.* 111 (3 July 2013), p. 030501. DOI: 10.1103/PhysRevLett.111.030501. URL: http://link.aps.org/doi/10.1103/PhysRevLett.111.030501.

[138] V. Caprara Vivoli et al. "Challenging preconceptions about Bell tests with photon pairs". In: *Phys. Rev. A* 91 (1 Jan. 2015), p. 012107. DOI: 10.1103/PhysRevA.91.012107. URL: http://link.aps.org/doi/10.1103/PhysRevA.91.012107.

[139] Alejandro Máttar et al. "Optimal randomness generation from optical Bell experiments". In: *New Journal of Physics* 17.2 (2015), p. 022003. URL: http://stacks.iop.org/1367-2630/17/i=2/a=022003.

[140] Stefano Pironio. "Lifting Bell inequalities". In: *Journal of Mathematical Physics* 46.6, 062112 (2005). DOI: http://dx.doi.org/10.1063/1.1928727. URL: http://scitation.aip.org/content/aip/journal/jmp/46/6/10.1063/1.1928727.

[141] Richard D. Gill. "Statistics, Causality and Bells Theorem". In: *Statist. Sci.* 29.4 (Nov. 2014), pp. 512–528. DOI: 10.1214/14-STS490. URL: http://dx.doi.org/10.1214/14-STS490.

[142] Daniel Collins and Nicolas Gisin. "A relevant two qubit Bell inequality inequivalent to the CHSH inequality". In: *Journal of Physics A: Mathematical and General* 37.5 (2004), p. 1775. URL: http://stacks.iop.org/0305-4470/37/i=5/a=021.

[143] Marco Tomamichel. "A framework for non-asymptotic quantum information theory". In: *arXiv preprint arXiv:1203.2142* (2012).