On constructions of quantum-secure device-independent randomness expansion protocols

By

Peter Johnson Brown

Doctor of Philosophy

University of York
Mathematics

May 2019

Abstract

Device-independent randomness expansion protocols aim to expand a short uniformly random string into a much longer one whilst guaranteeing that their output is truly random. They are device-independent in the sense that this guarantee does not dependent on the specifics of an implementation. Rather, through the observation of nonlocal correlations we can conclude that the outputs generated are necessarily random. This thesis reports a general method for constructing these protocols and evaluating their security. Using this method, we then construct several explicit protocols and analyse their performance on noisy qubit systems. With a view towards near-future quantum technologies, we also investigate whether randomness expansion is possible using current nonlocality experiments. We find that, by combining the recent theoretical and experimental advances, it is indeed now possible to reliably and securely expand randomness.

List of Contents

List of Figures

ACKNOWLEDGEMENTS

First and foremost, I must thank my supervisor Roger Colbeck for his support and guidance throughout the last four years. I have learnt a great deal and had a lot of fun in the process. Thanks Roger!

I would also like to thank my other collaborators Chris, Eleni and Sammy for many interesting discussions. On the topic of research, I would also like to thank all of the members (past and present) of the quantum information group. I thoroughly enjoyed our weekly seminars, dinners and discussions. To my office/house/PhD mates, I am incredibly grateful for the many fond memories and good times we've shared over the last few years! With sadness but also great appreciation, I must acknowledge the late Paul Busch, without whom I would likely never have pursued a PhD.

To my examiners, Omar Fawzi and Lluis Masanes, thank you for taking the time to read my thesis.

To my friends and family back home, thanks for the unconditional support and for always being there. Finally, a very special thanks goes to Mirjam - danke schön!

AUTHOR'S DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author. This work has been carried out under the supervision of Dr. Roger Colbeck and has not previously been presented for an award at this, or any other, university. Chapters 3 and 4 are largely based upon two research works that are listed below. All sources are acknowledged and have been listed in the Bibliography.

Related research works

(1) P. J. Brown, S. Ragy and R. Colbeck, *A framework for quantum-secure device-independent randomness expansion,* arXiv:1810.13346 (2018), (submitted to journal).

(2) W.-Z. Liu, M.-H. Li, S. Ragy, S.-R. Zhao, B. Bai, Y. Liu, P. J. Brown, J. Zhang, R. Colbeck, J. Fan, Q. Zhang and J.-W, Pan, *Experimental demonstration of device-independent randomness expansion* (in preparation).

CHAPTER 1

INTRODUCTION AND SYNOPSIS

Classical mechanics offers a deterministic explanation of the macroscopic world. If one is able to precisely model a classical system then all future observations are determined. Still, if a fair[1] coin is tossed behind an agent's back then, without additional information, the agent will struggle to correctly predict the outcome more than 50% of the time. Therefore, even though classical mechanics does not prescribe random events, a lack of predictability can still arise from sufficient ignorance. In stark contrast to this, quantum mechanics asserts that randomness is a fundamental feature of nature. In general, for quantum mechanical experiments, even complete knowledge of the system does not allow one to perfectly predict its outcomes. The existence of truly random events has profound consequences for applications like cryptography, where the unpredictability of a random sequence can be imprinted onto data which we wish to keep secret. By using a source of true randomness, we can ensure that even the most powerful adversaries are not privy to our secrets.

A demand for randomness was present long before Schrödinger's cat, with the earliest evidence of dice dating back several millennia. Ancient Athenians used lotteries as a means of electing officials [1] and gambling has been a prevalent feature within many historical cultures. A desire to better understand gambling and games of chance, drove 17^{th} century mathematicians towards a development of probability theory [2]. Then, as the disciplines of probability and statistics matured, scientific applications of randomness began to emerge. For example, random sampling is a key component of hypothesis testing and Monte Carlo simulations. More recently, owing to the seminal work of Shannon [3], randomness has found a plethora of applications to information processing and cryptography.

In cryptography one seeks to describe the ability of agents to complete some task whilst

[1] Here, fair means not purposefully weighted, i.e. the coin's density is approximately uniform.

in the presence of an adversary. For example, in a data encryption scheme an agent applies a randomly selected reversible transformation to some data such that no adversary is able to deduce the contents of the original data from the transformed data. However, for this to be possible, we require an additional feature from our source of randomness, namely we require a notion of *privacy*. That is, the output of a source of *private randomness* should be unpredictable from the perspective of all parties – including any would-be adversaries. This requirement presents a problem with building a random number generator suitable for cryptography: how can one ever ensure privacy?

When working with a classical source of randomness then evidently our privacy must come from some limitations placed on an adversary. Indeed, so-called cryptographically-secure pseudorandom number generators derive privacy from a computational perspective. Roughly, these are deterministic algorithms that when fed a short random seed produce a sequence of numbers that is computationally difficult to distinguish from a source of truly random numbers. These pseudorandom sources are by far the most commonly used sources, a consequence of their speed and low implementation costs (they do not require specialist hardware). However, popularity is not necessarily a good measure of security. In particular, the derivation of security from a computational perspective places privacy at the mercy of technological progress. The prime example of this is integer factoring [4], a task widely believed to be computationally difficult[2] was found to admit a polynomial time algorithm on a quantum computer [5].

Still, even if we base our randomness generation on a computational assumption which is valid, how can a user of such a source verify that an adversary does not have additional information that would render its outputs predictable? Without the technical expertise required to verify the source's construction, one is left to trust the manufacturer's claim of security. However, trust is not always warranted and this assumption could leave a user at the peril of incompetent or malicious manufacturers. It should be noted that alleged backdoors in standardised 'cryptographically-secure' random number generators have been reported [6].

Fortunately, the nonlocal characteristic of quantum theory provides a means to addressing the problem of generating certifiably private randomness. Entangled quantum systems can exhibit correlations between distant parties that are necessarily random (and private). Thus, through the observation of these correlations one is able to verify a source of randomness that is a priori private. Pioneered by the insights of [7, 8], this connection has developed into what is now known as device-independent quantum cryptography: the study of cryptographic procedures whose security can be established independently of the internal workings of any devices involved. The cryptographic primitives that have been tackled in the device independent setting are numerous and ever growing. Much of the early work [9, 10] focussed on quantum key distribution where two parties look to establish a

[2] A problem is computationally difficult if it cannot be solved using a polynomial time algorithm.

secret key between two distant locations. Other primitives that have been analysed include: randomness amplification [11, 12], the conversion of a source of randomness containing dependencies into a source of independent random bits; self-testing [13], certifying (up to local transformations) the internal quantum state and measurements present within some system; and bit-commitment [14, 15], a scenario wherein one party sends (commits) an encoded value to another which, at some point later, is revealed.

In this thesis, we focus on the task of randomness expansion. A procedure wherein one assumes access to a short private source of randomness and attempts to use it to generate a much larger (still private) source. Randomness expansion was originally proposed in [16,17] with further development and experimental testing following shortly after [18]. Subsequent work provided rigorous security proofs against classical adversaries [19, 20]. Security against quantum adversaries—who may share entanglement with the internal state of the device—came later [21–23], progressively increasing in noise-tolerance and generality, with the recently introduced entropy accumulation theorem (EAT) [24, 25], on which this work is based, providing asymptotically optimal rates [26, 27]. A new proof technique which is also asymptotically optimal has recently appeared [28].

The EAT has been applied to several cryptographic tasks [26, 27, 29–31]. All of these applications rely, at their core, on the CHSH test of nonlocality [32] or close variants thereof.[3] Special properties of the CHSH test are able to greatly simplify the analysis (for example, it can be shown that it is sufficient to consider the untrusted devices sharing qubit systems) [10]. However, these techniques cannot be directly generalized to the vast majority of other nonlocality tests. With respect for the ethos of device-independence, we should not assume that we can modify the untrusted devices to better suit the pre-existing protocols. Instead, in order to maximise randomness throughput, we should look to tailor the protocols to the devices under consideration.

In [26], it was suggested that one could look to use the device-independent guessing probability (DIGP) [37–39] in conjunction with the semidefinite hierarchy [40, 41] to obtain computational constructions of particular randomness bounding functions required by the EAT. In this thesis we detail precisely such a method, allowing us to apply the EAT to a wide range of nonlocality tests beyond CHSH (including the possibility of looking at multiple non-locality tests simultaneously and the inclusion of additional parties). We present this result in the form of a template randomness expansion protocol with security statements that can be evaluated numerically. Moreover, as this construction is both computationally efficient and robust, we are able to iteratively fine-tune the template protocol to best fit a given scenario. We apply the result to several different tests of nonlocality and compare their randomness expansion rates on entangled qubit pairs subject to noisy detectors. For each protocol considered we are able to generate close to two bits of randomness per

[3] In [29] the authors use a multipartite generalisation of CHSH known as an MABK inequality [33–35] and in [31] the authors use the tilted CHSH inequality [36].

entangled qubit pair (in the low noise regime).

In order to facilitate the understanding of this thesis, the relevant mathematical preliminaries have been collated to form Chapter 2. The expert reader may wish to skip this. The remaining chapters contain the main results of the thesis. Let us briefly summarise them here.

Chapter 3: In this chapter we introduce and prove security of a template randomness expansion protocol which can be adapted to the requirements of a user. We achieve this by establishing a connection between two powerful theoretical tools, the semidefinite hierarchy [40, 41], which is used to bound the device-independent guessing probability (DIGP) [37–39], and the entropy accumulation theorem EAT [24, 25]. We also provide a full analysis of the cost of seeding the protocol. In addition, a python package implementing the construction is provided [42]. This chapter is based on [43].

Chapter 4: In this chapter we apply our technique to several example protocols. In particular, we look at randomness expansion using the complete empirical distribution as well as a simple extension of the CHSH protocol, showing noise-tolerant rates of up to two bits per entangled qubit pair, secure against quantum adversaries. We also compare the achievable rates for these protocols to the protocol presented in [26] which is based upon a direct von Neumann entropy bound. Our comparison demonstrates that some of the protocols from the framework are capable of achieving higher rates than the protocol of [26], in both the low and high noise regimes. We conclude with a short analysis on the feasibility of randomness expansion with current experimental technologies. This analysis forms the theoretical basis of an upcoming manuscript [44].

We conclude in Chapter 5 with a discussion of the results presented and elaborate on some possible directions of further research.

CHAPTER 2

PRELIMINARIES

This chapter introduces the relevant mathematical tools required for understanding the remainder of this thesis. We begin by establishing some of the more general notation before providing a quick introduction to quantum information theory in Sec. 2.1. In Sec. 2.2 we define the relevant entropy measures and following this, in Sec. 2.3, we introduce several large deviation bounds including the entropy accumulation theorem. We then overview Bell nonlocality and device-independence in Sec. 2.5 and Sec. 2.6. Finally, in Sec. 2.7 we finish the preliminary material with an overview of semidefinite programming and its application to approximating the set of quantum correlations.

Notational conventions

Throughout this work, the calligraphic symbols $\mathcal{A}$, $\mathcal{B}$, $\mathcal{X}$ and $\mathcal{Y}$ denote finite sets (alphabets). We will use the notational shorthand $\mathcal{AB}$ to denote the Cartesian product alphabet $\mathcal{A}\times\mathcal{B}$. Given two random variables X and X', taking values in some common alphabet $\mathcal{X}$, the *statistical distance* between X and X', is defined as

$$\Delta(X,X') := \frac{1}{2}\sum_{x\in\mathcal{X}} |p_X(x) - p_{X'}(x)|. \tag{2.0.1}$$

We make frequent use of the bijective mapping between distributions and a subset of vectors in some real vector space. More specifically, after choosing some orthonormal basis $\{\boldsymbol{e}_{abxy}\}$ of $\mathbb{R}^{|\mathcal{ABXY}|}$ we may identify some distribution $p:\mathcal{ABXY}\to[0,1]$ on these alphabets with the vector $\boldsymbol{p} = \sum_{abxy} p(a,b,x,y)\boldsymbol{e}_{abxy}$. We refer to an element of our vector using the notation $p(a,b,x,y) = \boldsymbol{p}\cdot\boldsymbol{e}_{abxy}$. Given some subset of our indexing alphabets, $\mathcal{C}\subseteq\mathcal{ABXY}$, a *restriction of* $\boldsymbol{p}$ *to* $\mathcal{C}$, denoted $\boldsymbol{p}(\mathcal{C})$, is the result of applying the map $\Pi_{\mathcal{C}}:\mathbb{R}^{|\mathcal{ABXY}|}\to\mathbb{R}^{|\mathcal{C}|}$

defined by the action on the basis vectors as

$$\Pi_{\mathcal{C}} : \boldsymbol{e}_{abxy} \mapsto \begin{cases} \boldsymbol{e}_{abxy} & \text{if } (a,b,x,y) \in \mathcal{C} \\ \boldsymbol{0} & \text{otherwise.} \end{cases}$$

Given an alphabet $\mathcal{C}$ and a sequence $C_1^n = (c_i)_{i=1}^n$, with $c_i \in \mathcal{C}$ for each $i = 1,\ldots,n$, we denote the *frequency distribution* induced by C_1^n by

$$F_{C_1^n}(x) = \frac{\sum_{i=1}^n \delta_{xc_i}}{n}, \tag{2.0.2}$$

where δ_{ab} is the Kronecker delta on the set $\mathcal{C}$. If C_1^n is a sequence of random variables then we say that C_1^n is *i.i.d.* if all the random variables in the sequence are independent and identically distributed. For some *event* $\Omega \subseteq \mathcal{C}$, we write $\mathbb{P}[\Omega]$ to denote the probability that the event occurs. We denote the set of all probability distributions on the set $\mathcal{C}$ by $\mathfrak{P}_{\mathcal{C}}$. We shall use a tilde to denote subnormalisation, e.g. $\tilde{\boldsymbol{p}} \in \tilde{\mathfrak{P}}_{\mathcal{C}}$ denotes a possibly non-normalised distribution over $\mathcal{C}$, $\sum_{c\in\mathcal{C}} p(c) \leq 1$. In certain contexts we may refer to probability distributions as *behaviours* or *strategies*.

For a linear operator M on some Hilbert space $\mathcal{H}$, we define the *trace* of M to be $\mathrm{Tr}[M] = \sum_x \langle x|M|x\rangle$ where $\{|x\rangle\}$ is any orthonormal basis of $\mathcal{H}$. Furthermore, for a linear operator N on the tensor product of two Hilbert spaces $\mathcal{H}_A \otimes \mathcal{H}_B$, we define the *partial trace* over system A as $\mathrm{Tr}_A[N] = \sum_x (\langle x| \otimes \mathbb{1}_B) N (|x\rangle \otimes \mathbb{1}_B)$, where $\{|x\rangle\}_x$ is an orthonormal basis of $\mathcal{H}_A$. We refer to an operator $M \in \mathcal{L}(\mathcal{H})$ as *positive semidefinite* if $\langle\psi|M|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$. We denote positive semidefiniteness by $M \succeq 0$. This relation also induces a partial order on the space of linear operators (Loewner order), for $M, N \in \mathcal{L}(\mathcal{H})$ we say $M \succeq N$ if $M - N \succeq 0$.

We write the natural logarithm as $\ln(\cdot)$ and the logarithm base 2 as $\log(\cdot)$. The function

$$\mathrm{sgn}(x) := \begin{cases} 0 & \text{for } x = 0 \\ \frac{x}{|x|} & \text{otherwise} \end{cases} \tag{2.0.3}$$

is the sign function for $x \in \mathbb{R}$. We will also make use of the notation $[n] := \{1,2,\ldots,n\}$ for $n \in \mathbb{N}$.

2.1 Quantum information

Quantum information theory is the study of information processing tasks on quantum-mechanical systems. The purpose of this section is to define precisely what we mean by a *quantum system* and the physical laws that constrain it. We shall motivate these definitions from an operational viewpoint, introducing concepts from the perspective of *agents* (or experimentalists), typically referred to as Alice and Bob, who are capable of interacting with these systems. To begin with, we give an abstract definition of a *physical system*.

Definition 2.1. A *physical system* is a collection of objects $(\mathcal{S},\mathcal{T},\mathcal{M})$ that are defined as follows.

- $\mathcal{S}$ – *States*: Possible values that can be attributed to the internal degrees of freedom of the system.
- $\mathcal{T}$ – *Transformations*: A collection of mappings $T : \mathcal{S} \to \mathcal{S}'$ from the system's current state space to another.
- $\mathcal{M}$ – *Measurements*: A collection of mappings $M : \mathcal{S} \to \mathcal{S}' \times \mathcal{X}$, where $\mathcal{X}$ is some finite set labelling the possible outcomes of the measurements. Measurements probe the internal state of the system, returning some outcome $x \in \mathcal{X}$ and transforming the system's state.

Remark 2.1. When dealing with multiple physical systems it will be useful to give them labels to help with distinguishing. For example, we may refer to systems A and B that have state spaces $\mathcal{S}(A)$ and $\mathcal{S}(B)$ respectively. As transformations may alter the state space of a system (effectively defining a new system), we can also use these labels to keep track of a system at different times, i.e. A_1 may be the initial system then after applying a transformation or measurement we have a system A_2. Systems will be labelled using upper-case Roman characters.

Let us now introduce the states, transformations and measurements that constitute a quantum system.

Quantum states

A *quantum state* is a trace-one positive-semidefinite operator acting on some Hilbert space $\mathcal{H}$. Unless otherwise stated, a Hilbert space is assumed to be finite dimensional. We denote the set of all quantum states on the Hilbert space $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$. We say a state $\rho \in \mathcal{S}(\mathcal{H})$ is *pure* if it satisfies $\mathrm{Tr}\left[\rho^2\right] = 1$. After identifying some distinguished orthonormal basis $\{|x\rangle\}_x$ of $\mathcal{H}$ we refer to a state as *classical* if it is diagonal in this basis, i.e., $\rho_{\text{classical}} = \sum_x p(x)|x\rangle\langle x|$. Note that the unit trace condition, combined with the positive semidefinite property ensures that the vector $\boldsymbol{p} = (p(x))_x$ is a probability distribution. Representing random variables as quantum states allows for us to model standard (discrete) probability theory using the same language as quantum systems, hence the terminology 'classical'.

The joint state space of several quantum systems is the state space of the tensor products of the individual Hilbert spaces. More precisely, if we have a collection of $n \in \mathbb{N}$ quantum systems with the state space of the i^{th} quantum system being $\mathcal{S}(\mathcal{H}_i)$, then the state space of their joint system is $\mathcal{S}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n)$. We refer to a state $\rho_{\text{sep}} \in \mathcal{S}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n)$ as *separable* if it can be written as a convex combination of tensor products of states from the individual subsystems, that is, $\rho_{\text{sep}} = \sum_i \lambda_i \rho_{i,1} \otimes \cdots \otimes \rho_{i,n}$ with $\sum_i \lambda_i = 1$, $\lambda_i \geq 0$ and $\rho_{i,j} \in \mathcal{S}(\mathcal{H}_j)$ for every (i, j). A state which is not separable is called *entangled*.

A particular class of states which shall play an important role in the forthcoming work are so-called *classical-quantum* (cq) states. We say that a state $\rho_{\text{cq}} \in \mathcal{S}(XE)$ is a cq-state if it

takes the following form $\rho_{\text{cq}} = \sum_x p(x)|x\rangle\langle x| \otimes \rho_E^x$, where $\{|x\rangle\}_x$ is the distinguished 'classical' basis of $\mathcal{H}_X$ and $\rho_E^x \in \mathcal{S}(E)$ for each x.[1] Letting $\Omega \subseteq \mathcal{X}$ be an event on the alphabet $\mathcal{X}$, we define the *conditional state* (conditioned on the event Ω) by

$$\rho_{XE|\Omega} = \frac{1}{\mathbb{P}[\Omega]} \sum_{x \in \Omega} p(x)|x\rangle\langle x| \otimes \rho_E^x. \tag{2.1.1}$$

We denote the identity operator of a system E by $\mathbb{1}_E$.

Quantum channels

The set of allowed transformations (channels) is the set of linear mappings that preserve the state spaces. Precisely, these are any linear mappings $\Lambda : \mathcal{S}(A) \to \mathcal{S}(B)$ which satisfy the following two properties.

1. Λ is trace-preserving: for any $\rho_A \in \mathcal{S}(A)$ we have $\text{Tr}\left[\rho_A\right] = \text{Tr}\left[\Lambda(\rho_A)\right]$.

2. Λ is completely-positive: for any $\rho_{AC} \in \mathcal{S}(AC)$, if $\rho_{AC} \succeq 0$ then $(\Lambda \otimes \mathcal{I}_C)(\rho_{AC}) \succeq 0$.

These two conditions ensure that all quantum states are mapped to quantum states. This includes situations when we apply a transformation to only part of a system. Hence, we require not just positivity but also complete-positivity. We denote the set of all quantum channels between systems A and B by $\mathcal{T}(A,B)$ and we use the shorthand $\mathcal{T}(A)$ to denote $\mathcal{T}(A,A)$.

Quantum measurements

A measurement M with outcomes in the set $\mathcal{X}$ is described by a collection of positive-semidefinite operators $M = \{M_x\}_{x \in \mathcal{X}}$ acting on some Hilbert space $\mathcal{H}$. In addition to being positive, these operators are required to satisfy $\sum_{x \in \mathcal{X}} M_x = \mathbb{1}$. Applying the measurement M to a system in the state ρ, we receive the outcome $x \in \mathcal{X}$ with probability $p(x) = \text{Tr}\left[\rho M_x\right]$. We refer to such a measurement as a *positive operator valued measure* (POVM). If in addition the measurement operators are also projectors, i.e. $M_x^2 = M_x$ for all $x \in \mathcal{X}$, we refer to this as a *projection valued measure* (PVM). We denote the set of all measurements for a Hilbert space $\mathcal{H}_A$ by $\mathcal{M}(A)$.

The state transformation accompanying the measurement can be specified by a collection of quantum channels $\{\Lambda_x\}_{x \in \mathcal{X}}$, where Λ_x is applied to the state upon receiving the measurement outcome x. The exact structure of the channels associated with a measurement will depend upon the context. For example, physical constraints may dictate that a system ceases to exist post-measurement. In such a case we can treat the Λ_x as trivial channels, i.e. $\Lambda_x(\rho) = 1$ for all $x \in \mathcal{X}$ and $\rho \in \mathcal{S}$.

[1] Classical is a contextual concept: the spectral theorem tells us that for any state there exists a basis in which it is 'classical'. Rather, given some basis (say, defined by some measurement of interest) we can refer to states which are classical in this basis.

Definition 2.2. A *quantum system* A is a physical system with a state space $\mathcal{S}(A)$, the allowed transformations are quantum channels with domain $\mathcal{S}(A)$ and the allowed measurements are $\mathcal{M}(A)$ defined for some Hilbert space $\mathcal{H}_A$.

The *trace norm* specifies a norm on the space of linear operators acting on some Hilbert space. Let $\mathcal{H}, \mathcal{H}'$ be two Hilbert spaces and let $\mathcal{L}(\mathcal{H}, \mathcal{H}')$ be the set of linear operators from $\mathcal{H}$ to $\mathcal{H}'$. Then for any $A \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$, the trace norm of A is defined as

$$\|A\|_1 = \mathrm{Tr}\left[\sqrt{A^\dagger A}\right]. \tag{2.1.2}$$

From this, we also define the *trace distance* between two operators $A, B \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ as $\|A - B\|_1$.[2]

Remark 2.2. The trace distance has a useful operational interpretation as a quantity characterising the distinguishability of two quantum states [45]. Let $\lambda \in [0,1]$ and $\rho_0, \rho_1 \in \mathcal{S}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. Suppose that Alice sends either ρ_0 or ρ_1 to Bob with probability λ and $1-\lambda$ respectively. Upon receiving the state Bob is allowed to perform a measurement and subsequently guess which state he was sent.

The trace distance between the two states characterises this optimal probability of guessing exactly. Precisely, the maximum probability with which Bob can guess correctly is given by

$$\frac{1}{2} + \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1.$$

Furthermore, the measurement that achieves this is projective [46].

2.2 Entropies

In the context of this thesis, 'entropy' is semantically equivalent to 'randomness' and we will often use the two terms interchangeably. However, by taking such a linguistic stance we must remain vigilant of how our definition of randomness is related to the choice of entropy measure. In the following we precisely pin down our definition of randomness and its relation to entropy, drawing inspiration from the operational interpretations of the entropy measures that we introduce.

2.2.1 Classical entropy measures

Entropy, in the information-theoretic setting, was introduced by Shannon in his seminal work [3]. Guided by several information-theoretic questions, his strive to characterise a useful measure of information led him to define his now eponymous entropy.

[2] For classical states this definition coincides with the statistical distance (cf. (2.0.1)) up to a factor of $\frac{1}{2}$.

Definition 2.3. Let X be a random variable taking values in $\mathcal{X}$ with a probability distribution $\boldsymbol{p}$. The *Shannon entropy* of X is

$$H(X) := -\sum_{x\in\mathcal{X}} p(x)\log p(x), \tag{2.2.1}$$

where we also define $0\log 0 := 0$.[3]

Remark 2.3. As noted by Shannon, the entropy is defined independently of any meaning of X, depending only on the underlying distribution. Despite this, we use the standard notation $H(X)$ instead of $H(\boldsymbol{p})$.

Another relevant quantity is that of the *self-information* (or *surprisal*) $S : \mathcal{X} \to \mathbb{R}\cup\{\infty\}$, which is defined as

$$S(x) := -\log p(x). \tag{2.2.2}$$

It is related to the Shannon entropy by $H(X) = \mathbb{E}[S]$. One interpretation of the self-information is that of a function characterising the amount of surprise experienced by a rational agent upon observing some outcome $x \in \mathcal{X}$. That is, $-\log(p)$ is a monotonically decreasing function on the unit interval, with $-\log(1) = 0$ and $\lim_{p\to 0_+} -\log(p) = \infty$: events that occur with certainty produce no-surprise whereas impossible events are associated with infinite surprise. Moreover, for independent random variables the surprisal is additive. From this perspective, we can view the Shannon entropy as a measure of the expected surprisal experienced by the agent.

Several years later, in [47] Rényi reviewed Shannon's axioms of a 'good' entropy measure and relaxed what he referred to as *the mean value property* of entropy. This led him to define a one-parameter family of entropies which we refer to as the *Rényi entropies*.

Definition 2.4. Let X be a random variable taking values in $\mathcal{X}$ with a probability distribution $\boldsymbol{p}$ and let $\alpha \in [0,1)\cup(1,\infty)$. Then, the *$\alpha$-Rényi entropy* of X is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha}\log\sum_{x\in\mathcal{X}} p^\alpha(x). \tag{2.2.3}$$

Additionally, we define the limiting cases,

$$H_1(X) := H(X) \tag{2.2.4}$$

and

$$H_\infty(X) := \min_{x\in\mathcal{X}} -\log p(x). \tag{2.2.5}$$

Remark 2.4. We refer to the measure H_∞ as the *min-entropy*, denoting it by $H_{\min}$.

The Shannon entropy may differ substantially from other Rényi entropies as the follow example demonstrates.

[3] This choice follows from the identity $\lim_{x\to 0^+} x\log x = 0$.

Example 2.1. Let X be a random bit string of length m, $\mathcal{X} = \{0,1\}^m$, such that

$$\mathbb{P}[X = b] = \begin{cases} 1-\delta & \text{if } b = 0^m \\ \frac{\delta}{2^m - 1} & \text{otherwise} \end{cases} \tag{2.2.6}$$

for some $0 < \delta \leq 1/2$. For this distribution we have

$$\begin{aligned} H(X) &= h(\delta) + \delta \log(2^m - 1) \\ &\approx h(\delta) + \delta m \end{aligned}$$

and

$$H_{\min}(X) = -\log(1-\delta)$$

where $h(\delta) := -\delta\log(\delta) - (1-\delta)\log(1-\delta)$ is the binary entropy function. Here, we see that the Shannon entropy grows linearly in the number of bits whereas the min-entropy stays at a constant value. Furthermore, setting $\delta = 1/\sqrt{m}$ and taking the limit $m \to \infty$, we have $H(X) \to \infty$ whereas $H_{\min} \to 0$.

2.2.2 Quantum entropy measures

In the following we generalise the Shannon entropy and min-entropy to quantum systems. Let $\rho \in \mathcal{S}(A)$ be a quantum state, the *von Neumann entropy* of ρ is

$$H(A)_\rho := -\operatorname{Tr}\left[\rho \log(\rho)\right]. \tag{2.2.7}$$

If ρ is a classical state, i.e. $\rho = \sum_x p(x)|x\rangle\langle x|$ for some orthonormal set of vectors $\{|x\rangle\}_x$, we recover the Shannon entropy of the distribution $\{p(x)\}_x$,

$$H(A)_\rho = -\sum_x p(x) \log p(x). \tag{2.2.8}$$

For a bipartite state $\rho_{AE} \in \mathcal{S}(AE)$, let $\rho_E = \operatorname{Tr}_A\left[\rho_{AE}\right]$ and define the conditional von Neumann entropy of system A given system E when the joint system is in state ρ_{AE} by

$$H(A|E)_\rho := H(AE)_\rho - H(E)_\rho\,. \tag{2.2.9}$$

In addition, for a tripartite system $\rho_{ABE} \in \mathcal{S}(ABE)$, the conditional mutual information between A and B given E is defined as

$$I(A:B|E)_\rho := H(A|BE)_\rho - H(A|E)_\rho\,. \tag{2.2.10}$$

We drop the state subscript whenever the state is clear from the context.

For cq-states it is useful to consider the conditional min-entropy [48] in its operational formulation [39]. Given a cq-state $\rho_{XE} = \sum_x p(x)|x\rangle\langle x| \otimes \rho_E^x$, the maximum probability with which an agent holding system E can guess the outcome of a measurement on X is

$$p_{\text{guess}}(X|E) := \sup_{\{M_x\}_x} \sum_x p(x) \operatorname{Tr}\left[M_x \rho_E^x\right], \tag{2.2.11}$$

where the supremum is taken over all POVMs $\{M_x\}_x$ on system E. Using this we can define the min-entropy of a classical system given quantum side information as

$$H_{\min}(X|E) := -\log\left(p_{\text{guess}}(X|E)\right). \tag{2.2.12}$$

Another quantity of importance is the *ϵ-smooth conditional min-entropy*. Given some $\epsilon \geq 0$ and $\rho_{XE} \in \mathcal{S}(XE)$, the ϵ-smooth min-entropy $H^{\epsilon}_{\min}$ is defined as the supremum of the min-entropy over all states ϵ-close to ρ_{XE},

$$H^{\epsilon}_{\min}(X|E)_{\rho} := \sup_{\rho' \in B_{\epsilon}(\rho)} H_{\min}(X|E)_{\rho'}, \tag{2.2.13}$$

where $B_{\epsilon}(\rho)$ is the ϵ-ball centred at ρ defined with respect to the purified trace distance (see [49]). The interested reader is referred to [50] for a comprehensive overview of smooth quantum entropies.

2.3 Large deviation bounds

Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables, assuming the expectation value exists let $\mu = \mathbb{E}[X_1]$ and define the partial sum $S_n := \frac{1}{n}\sum_{i=1}^{n} X_i$. The law of large numbers, in its weak and strong forms, establishes a connection between long-term empirical observations S_n and the expected observation μ. More specifically, the weak law of large numbers states that [51], for every $t > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left[|S_n - \mu| > t\right] = 0. \tag{2.3.1}$$

That is, in the limit of infinitely many observations, the probability that the empirical average S_n deviates from the mean tends to zero. Probabilistic bounds on deviations for finite n also exist and are known in the wider literature as *concentration inequalities*, see [52]. We now introduce two such bounds that will prove useful in the later analysis.

The first is commonly known as the Chernoff bound [53], although we source our formulation from [54].

Lemma 2.1 (Chernoff bound)**.** *Let X_i be independent binary random variables for $i = 1,\ldots,n$, $S = \sum_i X_i$ and $\mu = \mathbb{E}[S]$. Then for $0 \leq t \leq 1$*

$$\mathbb{P}\left[S \geq (1+t)\mu\right] \leq e^{-t^2\mu/3}$$
$$\mathbb{P}\left[S \leq (1-t)\mu\right] \leq e^{-t^2\mu/2}.$$

Corollary 2.1. *For $r \leq \mu$ we have $\mathbb{P}\left[\left|S-\mu\right| \geq r\right] \leq 2e^{-r^2/(3\mu)}$.*

In addition to this, we also make use of Hoeffding's inequality [55].

Lemma 2.2 (Hoeffding's inequality)**.** *Let X_i be independent random variables, such that $a_i \leq X_i \leq b_i$ with $a_i, b_i \in \mathbb{R}$ for $i = 1,\ldots,n$. In addition, let $S = \sum_i X_i$ and $\mu = \mathbb{E}[S]$. Then for $t > 0$*

$$\mathbb{P}\left[|S-\mu| \geq t\right] \leq 2e^{-\frac{2t^2}{\sum_i (b_i - a_i)^2}}.$$

2.3.1 The AEP and its entropic generalisations

A statement analogous to the weak law of large numbers can be made if we consider the average surprisal of our observations $I_n := -\frac{1}{n}\log p(X_1,\ldots,X_n)$. Note that due to the independence of the random variables we may rewrite the average surprisal as the sum of the individual surprisals $I_n = \frac{1}{n}\sum_{i=1}^{n} -\log(p(X_i))$. Writing $Y_i = -\log(p(X_i))$, we have $(Y_i)_{i\in\mathbb{N}}$ is a sequence of independent random variables with $\mathbb{E}[Y_1] = H(X_1)$. Applying the weak law of large numbers, we find that for all $t > 0$

$$\lim_{n\to\infty} \mathbb{P}\left[\left|-\frac{1}{n}\log p(X_1,\ldots,X_n) - H(X)\right| > t\right] = 0. \tag{2.3.2}$$

This result is known as the *asymptotic equipartition property* (AEP).

An equivalent entropic reformulation of the AEP can be made about the convergence of smooth Rényi entropies to the Shannon Entropy in the i.i.d. limit, see Chapter 6 in [56] for a complete discussion regarding this connection. In particular, for the smooth min-entropy we have

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \tfrac{1}{n} H_{\min}^{\epsilon}(X_1,\ldots,X_n) = H(X). \tag{2.3.3}$$

The advantage of considering the entropic form of the AEP is twofold. Firstly, entropies are operationally relevant quantities, bounds on their values imply bounds on operational problems. In addition, the entropic form allows us to avoid the problem of not having a well-defined generalisation of conditional probability due to the presence of incompatible joint events in quantum theory. Thus, a generalisation of (2.3.3) to quantum entropies in the presence of quantum side-information is possible [57]. The main result of [57] may be summarised as follows: for a quantum state $\rho \in \mathcal{S}(AB)$ consider the i.i.d. state $\rho^{\otimes n} := \bigotimes_{i=1}^{n}\rho$ on the system A^nB^n comprised of n copies of AB, then

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \tfrac{1}{n} H_{\min}^{\epsilon}(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_{\rho}. \tag{2.3.4}$$

Furthermore, for finite n the authors provide an explicit lower bound of the form

$$\tfrac{1}{n} H_{\min}^{\epsilon}(A^n|B^n)_{\rho^{\otimes n}} \geq H(A|B)_{\rho} - \tfrac{c}{\sqrt{n}}, \tag{2.3.5}$$

where c is a constant dependent on ϵ but independent of n. Lower bounds on the smooth min-entropy are of critical importance to later applications (c.f. Sec. 2.4).

Remark 2.5. The above bound (2.3.5) illustrates the power of entropy smoothing. Recall in Example 2.1 we saw that the Shannon entropy could be made to be arbitrarily larger than the min-entropy. The AEP states that for a long enough sequence of i.i.d. experiments we can account for this difference by smoothing. Intuitively, this can be seen as a consequence of typicality [51]. That is, for sufficiently large n almost all of the mass of $\boldsymbol{p}_{X_1^n}$ will be on the typical set $A_\delta^n := \{x_1^n \in \mathcal{X}^n \mid 2^{-n(H(X)+\delta)} \leq p(x_1^n) \leq 2^{-n(H(X)-\delta)}\}$. Therefore, one can find a distribution close to $\boldsymbol{p}_{X_1^n}$ such that $\frac{1}{n}H_{\min}(X_1^n) \gtrsim H(X) - \delta$.

2.3.2 The entropy accumulation theorem

The final bound in the previous subsection (2.3.5) is structured in a manner that is particularly appealing for our purposes. It lower bounds a global quantity of interest $H_{\min}^{\epsilon}(A^n|B^n)$ in terms of smaller, more easily computable quantities $H(A|B)_\rho$. Unfortunately, the bound is derived from the assumption of an i.i.d. structure to the sequence of experiments. This i.i.d. assumption in cryptography is overly restrictive. When characterising the possible actions of an adversary we would much rather be overly generous than overly restrictive.[4] Thus, we look for a bound similar to (2.3.5) which is derived from weaker assumptions. Fortunately, this is precisely what the entropy accumulation theorem (EAT) provides [24, 25].

The EAT relaxes the i.i.d. assumption by replacing it with a structural assumption which states that the random experiments are performed sequentially. That is, we conduct experiment one, then experiment two, then three and so on. The only additional a priori restriction that the sequential scenario imposes is that we should be able to explain any dependencies between two experiments by the information present at the earlier one. The remaining part of this section is dedicated to precisely defining this sequential interaction and stating the generalisation of (2.3.5) offered to us by the EAT. It should be noted here that we are not considering the EAT in its full generality, but a restricted version tailored to our application, e.g. some systems will be restricted to be classical.

In the sequential scenario our experiment begins with a bipartite system R_0E in a (possibly entangled) quantum state $\rho_{R_0E} \in \mathcal{S}(R_0E)$. We can think of the subsystem R_0 as the initial system in our laboratory, i.e. the system with which we shall be interacting sequentially. Whereas, system E refers to a system held by an adversary. During each interaction some classical systems $A_iX_iC_i$ will be produced along with a quantum system R_i. The sequential interaction is then described by the application of a sequence of quantum channels $(\mathcal{N}_i \otimes \mathcal{I}_E)_{i=1}^n$, where $\mathcal{N}_i : \mathcal{S}(R_{i-1}) \to \mathcal{S}(A_iX_iC_iR_i)$ which transform the system in the laboratory to produce the relevant experimental data. We refer to the channels used as *EAT channels* and the following definition gives their precise characterisation.

Definition 2.5 (EAT channels)**.** A set of *EAT channels* $\{\mathcal{N}_i\}_{i=1}^n$ is a collection of quantum channels $\mathcal{N}_i : \mathcal{S}(R_{i-1}) \to \mathcal{S}(A_iX_iC_iR_i)$ such that for each $i \in [n]$:

1. A_i, X_i and C_i are finite dimensional classical systems. Moreover, the state of C_i is a deterministic function of the states of A_i and X_i. R_i is an arbitrary quantum system.

2. For any initial state ρ_{R_0E}, the final state $\rho_{A_1^nX_1^nE} = \mathrm{Tr}_{C_1^nR_n}\left[((\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1) \otimes \mathcal{I}_E)\rho_{R_0E}\right]$ obeys the collection of conditional independence constraints $I(A_1^{i-1}{:}X_i|X_1^{i-1}E) = 0$.

Remark 2.6. The EAT channels formalise the notion of interaction within the protocol. The first constraint is a straightforward restriction on the nature of the information

[4] Furthermore, any innocent fluctuations in the experimental setup would render the i.i.d. assumption invalid.

present within the experiment. In the applications to come one can interpret these systems as follows: the systems (A_i) refer to outputs of measurements on the quantum systems (R_{i-1}) present in the laboratory at the beginning of the i^{th} interaction, we will use the EAT to bound the total entropy of these systems; the systems (X_i) correspond to any additional classical information entering the experiment during each interaction, e.g. a choice of measurement to make on the laboratory's system; finally (C_i) are systems that help evaluate the quality of the data produced during the interaction, e.g. the evaluation of some Bell-inequality for (A_i, X_i). We refer to the C_i systems as *scores*. Crucially, we will be able to evaluate our lower bound on the total entropy accumulated whilst conditioning on the observation of some event, i.e. $C_1^n \in \Omega$. For example, this could be the observation of a sufficiently large average score.

The independence constraints impose the idea that the outputs generated in the sequential interactions should only depend upon the information present up to that point in time. If these conditional independence conditions were abandoned then we could artificially construct these dependencies by choosing X_{i+1} based upon the value of some past output A_i. An explicit example showing the necessity of this condition for the general EAT statement is provided in the appendix of [24].

To fully state the EAT we require one additional object known as a *min-tradeoff function*. Such a function acts as a lower bound on the minimum von Neumann entropy accrued during a single interaction when we presuppose that the score will behave according to some distribution. Formally, it is defined as follows.

Definition 2.6 (Min-tradeoff function)**.** Let $\mathcal{N}_i$ be an EAT channel and let R' be a quantum system isomorphic to the system R_{i-1}. A *min-tradeoff function* for the channel $\mathcal{N}_i$ is an affine function $f_{\min} : \mathfrak{P}_C \to \mathbb{R}$ satisfying

$$f_{\min}(\boldsymbol{p}) \leq \inf_{\omega \in \Sigma_{\boldsymbol{p}}} H(A_i|X_iR')_{\mathcal{N}_i(\omega)} \tag{2.3.6}$$

where

$$\Sigma_{\boldsymbol{p}} := \left\{ \omega \in \mathcal{S}(R_{i-1}R') : (\mathcal{N}_i \otimes \mathcal{I}_{R'})(\omega)_{C_i} = \sum_c p(c)\,|c\rangle\langle c| \right\}. \tag{2.3.7}$$

When $\Sigma_{\boldsymbol{p}} = \emptyset$ the infimum is defined to be $+\infty$. For a given min-tradeoff function $f_{\min}$, we are also concerned about the following properties.

- Maximum over all distributions:

$$\text{Max}[f_{\min}] := \max_{\boldsymbol{p} \in \mathfrak{P}_C} f_{\min}(\boldsymbol{p}). \tag{2.3.8}$$

- Minimum over all distributions:

$$\text{Min}[f_{\min}] := \min_{\boldsymbol{p} \in \mathfrak{P}_C} f_{\min}(\boldsymbol{p}). \tag{2.3.9}$$

- Minimum over all Σ-compatible distributions:

$$\mathrm{Min}_{\Sigma}[f_{\min}] := \min_{\boldsymbol{p}:\Sigma_{\boldsymbol{p}}\neq\emptyset} f_{\min}(\boldsymbol{p}). \tag{2.3.10}$$

- Maximum variance over all Σ-compatible distributions:

$$\mathrm{Var}_{\Sigma}[f_{\min}] := \max_{\boldsymbol{p}:\Sigma_{\boldsymbol{p}}\neq\emptyset} \sum_{c} p(c)\left(f_{\min}(\boldsymbol{\delta}_c) - f_{\min}(\boldsymbol{p})\right)^2. \tag{2.3.11}$$

Remark 2.7. As a min-tradeoff function f is an affine function, we may decompose its action on a probability distribution $\boldsymbol{p}$ as $f(\boldsymbol{p}) = \sum_c p(c) f(\boldsymbol{\delta}_c)$.

The sequential interaction scenario imposes little structure on the experiment: we have a sequence of experiments obeying some statistical properties. At the end of the experiment we will have collected the statistics C_1^n, which gives rise to a frequency distribution over the set $\mathcal{C}$ which we denote by $F_{C_1^n}$. Loosely, the EAT states that the total entropy gained throughout the series of interactions should, with high probability, be close to the minimum entropy accumulated in a series of experiments whose scores are all i.i.d. according to $F_{C_1^n}$. The EAT was originally stated in [24], here we use the recently improved statement [25] wherein a better error dependence was established.

Theorem 2.2 (EAT)**.**
Let $(\mathcal{N}_i)_{i=1}^n$ be a collection of EAT channels and let $\rho_{A_1^n X_1^n C_1^n E} = \mathrm{Tr}_{R_n}\left[((\mathcal{N}_n \circ \cdots \circ \mathcal{N}_1)\otimes \mathcal{I}_E)\rho_{R_0 E}\right]$ be the output state after the sequential application of the channels $(\mathcal{N}_i \otimes \mathcal{I}_E)_i$ to some input state ρ_{R_0E}. Let $\Omega \subseteq \mathcal{C}^n$ be some event that occurs with probability p_Ω and let $\rho_{|\Omega}$ be the final state conditioned on Ω occurring. Finally let $\epsilon_s \in (0,1)$ and f_{min} be a min-tradeoff function for each $(\mathcal{N}_i)_i$. If for all $C_1^n \in \Omega$, with $\mathbb{P}\left[C_1^n\right] > 0$, we have $f_{min}(F_{C_1^n}) \geq t$ for some $t \in \mathbb{R}$ then for any $\beta \in (0,1)$ we have

$$H_{\min}^{\epsilon_s}(A_1^n|X_1^n E)_{\rho_{|\Omega}} > nt - n(\epsilon_V + \epsilon_K) - \epsilon_\Omega, \tag{2.3.12}$$

where

$$\epsilon_V := \frac{\beta \ln 2}{2}\left(\log\left(2|\mathcal{A}|^2+1\right) + \sqrt{\mathrm{Var}_{\Sigma}[f_{min}]+2}\right)^2, \tag{2.3.13}$$

$$\epsilon_K := \frac{\beta^2}{6(1-\beta)^3 \ln 2} 2^{\beta(\log|\mathcal{A}| + \mathrm{Max}[f_{min}] - \mathrm{Min}_{\Sigma}[f_{min}])} \ln^3\left(2^{\log|\mathcal{A}| + \mathrm{Max}[f_{min}] - \mathrm{Min}_{\Sigma}[f_{min}]} + e^2\right) \tag{2.3.14}$$

and

$$\epsilon_\Omega := \frac{1}{\beta}(1 - 2\log(p_\Omega \epsilon_s)). \tag{2.3.15}$$

2.4 Randomness extraction

Many applications require some source of uniform (or almost uniform) random bits. However, it may be the case that we only have access to a source of partially random bits. In

addition, these bits may not be independent of each other or, worse, they may be correlated with some other random source held by an adversary. Randomness extractors are procedures that allow us to process partially random sources into sources of independent and uniformly random bits.

The first extractor is attributed to von Neumann [58] who proposed a simple method for converting a source of biased but i.i.d. random bits into a uniformly random source.[5] More modern extractors began with technical results like the *leftover hash lemma* (LHL) [59]. The LHL states that if our partially random source of bits has min-entropy larger than $k > 0$ and we apply a function chosen randomly from a set of *pairwise-independent universal hash functions*[6] then the resulting bit string is approximately uniformly distributed over $\{0,1\}^k$. This result was later extended to the case where our random source may be correlated with some quantum system [60], proving that the results of quantum measurements could also be used as a source of extractable randomness. We now present the formal definition of a randomness extractor that will be used throughout the rest of this thesis. This definition is the combination of Lemma 3.5 from [61] and the quantum-proof randomness extractor definition presented in [62].

Definition 2.7 (Quantum-proof strong extractor)**.** We say that a function $R_{\text{ext}} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^r$ is a *quantum-proof* $(k, \epsilon_{ext} + 2\epsilon_s)$*-strong extractor*, if for all cq-states ρ_{XE} with $H_{\min}^{\epsilon_s}(X|E)_\rho \geq k$ for some $\epsilon_s > 0$ it maps $\rho_{XE} \otimes \tau_d$ to $\rho'_{R_{\text{ext}}(X,D)DE}$ where

$$\frac{1}{2}\| \rho'_{R_{\text{ext}}(X,D)DE} - \tau_r \otimes \tau_d \otimes \rho_E \|_1 \leq \epsilon_{\text{ext}} + 2\epsilon_s, \tag{2.4.1}$$

where τ_r is the maximally mixed state on a system of dimension 2^r.

Although in general the amount of randomness extracted will depend on the extractor, $H_{\min}^{\epsilon_s}(A_1^n B_1^n|E)$ provides an upper bound on the total number of ϵ_s-close to uniform bits that can be extracted from $A_1^n B_1^n$ and a well-chosen extractor will result in a final output bit-string of length $r \approx H_{\min}^{\epsilon_s}(A_1^n B_1^n|E)$. We denote any loss of entropy incurred by the extractor by $\ell_{\text{ext}} = k - r$. Entropy loss will differ between extractors but in general it will be some function of the extractor error, the seed length and the initial quantity of smooth min-entropy. The extractor literature is rich with explicit constructions, with many following Trevisan's framework [63]. For an in-depth overview of randomness extraction, we refer the reader to [64] and references therein.

[5] His idea is as follows: take the outputs of the source and group them into pairs. Each pair (b_1, b_2) takes one of four possible values $\{00, 01, 10, 11\}$ and, as the sequence was produced in an i.i.d. manner, $\mathbb{P}[(b_1, b_2) = (0,1)] = \mathbb{P}[(b_1, b_2) = (1,0)]$. If we discard all pairs taking values $(0,0)$ or $(1,1)$ we are left with a sequence of uniformly random i.i.d. binary-outcome events.

[6] Here, a hash function is just some map $h : \{0,1\}^n \to \{0,1\}^k$ where n is the number of bits produced by our partially random source. In addition, we say a collection of hash functions H is a *pairwise-independent universal family* if for all $x, y \in \{0,1\}^n$ with $x \neq y$ if we select $h \in H$ uniformly at random then we have $(h(x), h(y))$ is uniformly distributed over $\{0,1\}^{2k}$.

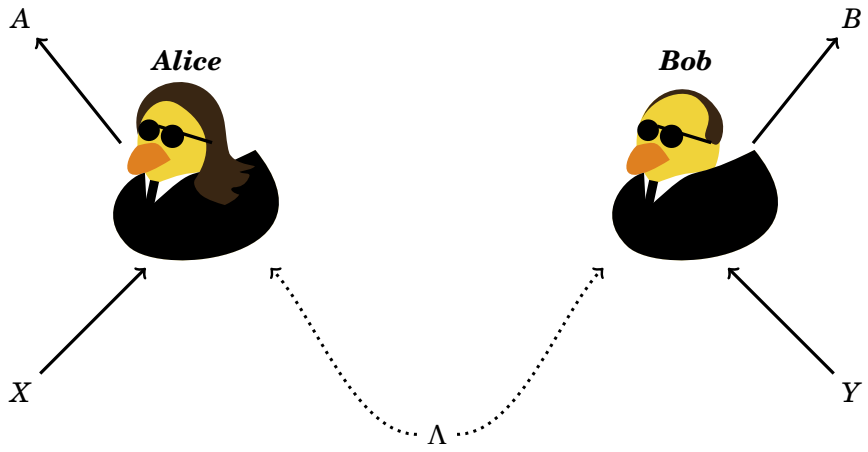


Figure 2.1: An artist's interpretation of two agents conducting a Bell-test.

Remark 2.8. By using a *strong* quantum-proof extractor, the output of the extractor will remain uncorrelated with the string used to seed it. Since the seed acts like a catalyst, we need not be overly concerned with the amount required. Furthermore, if available, it could just be acquired from a trusted public source immediately prior to extraction without compromising security.

2.5 Bell Nonlocality

Consider the following scenario (see Fig. 2.1) wherein two agents, Alice and Bob, are tasked with generating distant correlations. At the beginning of the experiment they will be separated and unable to communicate.[7] They will then each receive a symbol, Alice receives some $x \in \mathcal{X}$ and Bob receives some $y \in \mathcal{Y}$, chosen independently of the rest of the experiment. We denote the symbols which Alice and Bob receive by the random variables X and Y respectively. After receiving X and Y, they will each announce another symbol, A and B chosen from the sets $\mathcal{A}$ and $\mathcal{B}$ respectively. The outcome of a single run of the experiment, i.e. the tuple (A,B,X,Y), can be modelled by some conditional probability distribution $\boldsymbol{p}_{AB|XY}$.[8]

The separation of the agents enforces what are known as *no-signalling* constraints on the distribution $\boldsymbol{p}$,

$$\sum_a p(a,b|x,y) = p(b|y), \qquad \text{for each } b,x \text{ and } y, \tag{2.5.1}$$

$$\sum_b p(a,b|x,y) = p(a|x), \qquad \text{for each } a,x \text{ and } y. \tag{2.5.2}$$

[7] This assumption can be physically justified through spacelike separation of the two agents.

[8] The distribution over the inputs is fixed and cannot be influenced by the agents. Our attention is therefore focussed on the conditional distribution: we want to observe how the agents act upon receiving different inputs.

That is, the input received by one agent cannot influence the marginal distribution of the other agent.

Note that we allow the two agents, prior to being separated, to discuss how they will react upon receiving their inputs e.g. they could exchange some random bit string which would influence their choice of responses. What we are assuming here is that any dependencies between the outcomes A and B must be mediated by some underlying *hidden random variable* (HRV) Λ that was shared prior to the experiment. Mathematically, this corresponds to their probability distribution decomposing as

$$\begin{aligned} p(a,b|x,y) &= \sum_{\lambda\in\Lambda} q(\lambda)p(a|x,y,\lambda)p(b|x,y,\lambda) \\ &= \sum_{\lambda\in\Lambda} q(\lambda)p(a|x,\lambda)p(b|x,\lambda), \end{aligned} \tag{2.5.3}$$

where on the second line we used the assumption that A is independent of Y and B is independent of X.

Henceforth, a *behaviour* refers to a bipartite conditional probability distribution of the form $\{p(a,b|x,y)\}_{abxy}$ with $(a,b,x,y)\in\mathcal{ABXY}$ which satisfies the no-signalling constraints (2.5.1) and (2.5.2). We denote the set of *local behaviours*, i.e. those which can be decomposed as (2.5.3), by $\mathfrak{L}$. The set of local behaviours forms a bounded convex polytope in $\mathbb{R}^{|\mathcal{ABXY}|}$. Convex polytopes can be described in two equivalent ways: either as the convex hull of a set of extremal vertices or as the intersection of a collection of halfspaces, with the halfspaces being defined by the hyperplanes that lie across the facets of the polytope. For the polytope $\mathfrak{L}$, the extremal vertices are the deterministic behaviours [65], i.e. a behaviour of the form $p(a,b|x,y)=p(a|x)p(b|y)$ with $p(a|x),p(b|y)\in\{0,1\}$ for each $(a,b,x,y)\in\mathcal{ABXY}$.

Whilst the vertex-description of $\mathfrak{L}$ is simple to define, it is the halfspace description that has proven vastly more useful in practice.[9] Consider a hyperplane H that rests on a facet of $\mathfrak{L}$, it may be written as $\{\boldsymbol{p}\in\mathbb{R}^{|\mathcal{ABXY}|} \mid \sum_{a,b,x,y} s_{abxy}p(a,b|x,y)=c\}$ for some $s_{abxy},c\in\mathbb{R}$. This hyperplane defines the linear functional $S:\boldsymbol{p}\mapsto\sum_{a,b,x,y}s_{abxy}p(a,b|x,y)$ such that for $\boldsymbol{p}\in\mathfrak{L}$ we have $S(\boldsymbol{p})\leq c$.[10] This necessary criterion for membership of $\mathfrak{L}$ is known as a *Bell-inequality* and its usefulness comes from the fact that checking $S(\boldsymbol{p})>c$ is sufficient to conclude $\boldsymbol{p}\notin\mathfrak{L}$.

In his seminal work, [67], Bell extended the EPR argument [68] to conclude that the predictions of quantum theory could not always be described by the LHV model, i.e., quantum theory allows for probability distributions that cannot be decomposed in the manner depicted in (2.5.3). Such distributions are referred to as *nonlocal*. Since then, a plethora of work has been devoted to the study of nonlocality from both the foundational

[9] The process of converting a vertex description of a polytope to a hyperplane description is known as *facet enumeration*. Whilst polynomial time algorithms exist [66], a significant problem is that the number of vertices scales exponentially in the size of the input-output alphabets – the number of deterministic distributions is $|\mathcal{A}|^{|\mathcal{X}|}|\mathcal{B}|^{|\mathcal{Y}|}$.

[10] We will always have the our inequalities directed in this way. If required, we can make the replacements $s_{abxy}\mapsto -s_{abxy}$ and $c\mapsto -c$.

perspective as well as its implications for information processing tasks. We refer the reader to [69] and references therein for a broad overview of the topic. We refer to a behaviour $\boldsymbol{p}$ as *quantum*, denoted by $\boldsymbol{p} \in \mathfrak{Q}$, if there exists a Hilbert space $\mathcal{H}_{AB}$, a density operator $\rho \in \mathcal{S}(AB)$ and a collection of POVMs, $\{\{F_{a|x}\}_{a\in\mathcal{A}}\}_{x\in\mathcal{X}}$ and $\{\{G_{b|y}\}_{b\in\mathcal{B}}\}_{y\in\mathcal{Y}}$ such that $p(a,b|x,y) = \mathrm{Tr}\left[\rho(F_{a|x}\otimes G_{b|y})\right]$ for each $(a,b,x,y) \in \mathcal{AB}\mathcal{X}\mathcal{Y}$. Finally, the set of behaviours that are restricted only by the no-signalling constraints is denoted by $\mathfrak{N}$.

Example 2.2 (CHSH)**.** Let us consider the simplest non-trivial example of Bell-nonlocality, where $\mathcal{A} = \mathcal{B} = \mathcal{X} = \mathcal{Y} = \{0,1\}$. In this scenario there is a unique (up to relabelling) Bell inequality, *the CHSH inequality* [32]. The inequality takes the form

$$\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle \leq 2, \tag{2.5.4}$$

where $\langle A_xB_y\rangle := \sum_{ab}(-1)^{a\oplus b}p(a,b|x,y)$ for each $x,y \in \{0,1\}$. Its maximum violation for quantum systems, known as its Tsirelson bound [70], is $2\sqrt{2}$.

Remark 2.9. The crucial property of nonlocal correlations, that enables their application to cryptographic tasks, is the fact that they imply the existence of private randomness [9, 71]. More precisely, let $\boldsymbol{p}$ be a behaviour such that $\boldsymbol{p} \notin \mathfrak{L}$. Then, by the definition of $\mathfrak{L}$, there cannot exist a random variable Λ such that conditioned on knowing the value of Λ, A is a deterministic function of X and B is a deterministic function of Y. Thus, for some values of (x,y) we have $\max_{ab} p(a,b|x,y,\Lambda) < 1$ and therefore private randomness.

2.5.1 Nonlocal games

An alternative way of thinking about Bell-experiments is through the guise of a *nonlocal game*. From this viewpoint, a Bell-experiment is a cooperative game wherein two agents are separated and then a referee sends each agent a question to which they respond with an answer. Based upon their interaction, the two agents receive a *score* which is a function of their question and answers.

Definition 2.8. Let $\mathcal{A},\mathcal{B},\mathcal{X},\mathcal{Y}$ and $\mathcal{V}$ be finite sets. A (two-player) nonlocal game $\mathcal{G} = (\mu, V)$ (on $\mathcal{AB}\mathcal{X}\mathcal{Y}$) consists of a set of question pairs $(x,y) \in \mathcal{X}\mathcal{Y}$ chosen according to some probability distribution $\mu : \mathcal{X}\mathcal{Y} \to [0,1]$, a set of answer pairs $(a,b) \in \mathcal{AB}$ and a scoring function $V : \mathcal{AB}\mathcal{X}\mathcal{Y} \to \mathcal{V}$.

Remark 2.10. We will abuse notation and use the symbol $\mathcal{G}$ to refer to both the nonlocal game and the set of possible scores. I.e., we may refer to the agents receiving a score $c \in \mathcal{G}$. Furthermore, we denote the number of different scores by $|\mathcal{G}|$.

As before, the actions of our agents are modelled by some conditional probability distribution $\boldsymbol{p} \in \mathfrak{N}$, which we may refer to as a *strategy*. By playing according to the strategy

$\boldsymbol{p}$, the agents induce a frequency distribution $\boldsymbol{\omega}_{\mathcal{G}}$ over the set of possible scores. That is,

$$\omega_{\mathcal{G}}(c) = \sum_{abxy} \mu(x,y)p(a,b|x,y)\delta_{V(a,b,x,y),c} \tag{2.5.5}$$

for each $c \in \mathcal{G}$. We denote the set of possible frequency distributions achievable by the agents whilst playing according to quantum strategies by $\mathfrak{Q}_{\mathcal{G}}$.

Example 2.3 (CHSH game)**.** We may recast the CHSH inequality as a nonlocal game. In this game μ is the uniform distribution over $\mathcal{X}\mathcal{Y}$ and the scoring function is

$$V(a,b,x,y) = \begin{cases} 1 & \text{if } a \oplus b = xy \\ 0 & \text{otherwise.} \end{cases} \tag{2.5.6}$$

Interpreting a score of 1 as a win, the maximum probability with which the agents can win the CHSH game whilst using a local strategy is

$$\sup_{\boldsymbol{p}\in\mathfrak{L}} \omega(1) = \tfrac{3}{4}$$

and whilst playing according to a quantum behaviour is

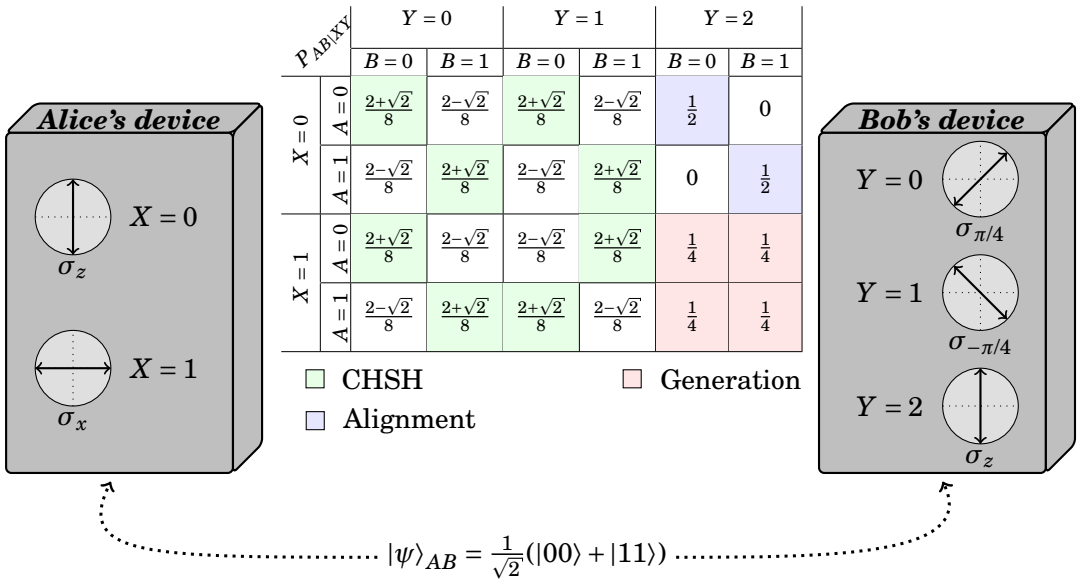
$$\sup_{\boldsymbol{p}\in\mathfrak{Q}} \omega(1) = \tfrac{1}{2} + \tfrac{\sqrt{2}}{4}.$$

We finish this section with another example of a nonlocal game, one which will be important for the analysis later in the thesis.

Example 2.4 (Extended CHSH game ($\mathcal{G}_{\text{CHSH}}$))**.** The *extended CHSH game* has appeared already in the device-independent literature [26, 72, 73]. It extends the standard CHSH game to include a correlation check between one of Alice's CHSH inputs and an additional input from Bob. It is defined by the question-answer sets $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1,2\}$ and $\mathcal{A} = \mathcal{B} = \{0,1\}$, the scoring set $\mathcal{V} = \{c_{\text{CHSH}}, c_{\text{align}}, 0\}$ and the scoring rule

$$V_{\text{CHSH}}(a,b,x,y) := \begin{cases} c_{\text{CHSH}} & \text{if } x \cdot y = a \oplus b \text{ and } y \neq 2 \\ c_{\text{align}} & \text{if } (x,y) = (0,2) \text{ and } a \oplus b = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.5.7}$$

The input distribution we consider is defined by $\mu_{\text{CHSH}}(x,y) = \frac{1}{8}$ for $(x,y) \in \{0,1\}^2$, $\mu_{\text{CHSH}}(0,2) = \frac{1}{2}$ and $\mu_{\text{CHSH}}(x,y) = 0$ otherwise. This game is equivalent to choosing to play either the CHSH game or the game corresponding to checking the alignment of the inputs $(0,2)$ uniformly at random and then proceeding with the chosen game. The frequency distribution then tells us the relative frequencies with which we win each game. The motivation behind $\mathcal{G}_{\text{CHSH}}$ can be understood by considering a schematic of an ideal implementation on a bipartite qubit system as given in Fig. 2.2. If we observe the maximum winning probability for the CHSH game, as well as perfect alignment for the inputs $(0,2)$, then the inputs $(\tilde{x},\tilde{y}) = (1,2)$ should produce two perfectly uniform bits.



$P_{AB\|XY}$		$Y=0$		$Y=1$		$Y=2$	
		$B=0$	$B=1$	$B=0$	$B=1$	$B=0$	$B=1$
$X=0$	$A=0$	$\frac{2+\sqrt{2}}{8}$	$\frac{2-\sqrt{2}}{8}$	$\frac{2+\sqrt{2}}{8}$	$\frac{2-\sqrt{2}}{8}$	$\frac{1}{2}$	0
	$A=1$	$\frac{2-\sqrt{2}}{8}$	$\frac{2+\sqrt{2}}{8}$	$\frac{2-\sqrt{2}}{8}$	$\frac{2+\sqrt{2}}{8}$	0	$\frac{1}{2}$
$X=1$	$A=0$	$\frac{2+\sqrt{2}}{8}$	$\frac{2-\sqrt{2}}{8}$	$\frac{2-\sqrt{2}}{8}$	$\frac{2+\sqrt{2}}{8}$	$\frac{1}{4}$	$\frac{1}{4}$
	$A=1$	$\frac{2-\sqrt{2}}{8}$	$\frac{2+\sqrt{2}}{8}$	$\frac{2+\sqrt{2}}{8}$	$\frac{2-\sqrt{2}}{8}$	$\frac{1}{4}$	$\frac{1}{4}$

Figure 2.2: A measurement schematic for a qubit implementation of $\mathcal{G}_{\text{CHSH}}$. Measurements are depicted in the x-z plane of the Bloch-sphere with $\sigma_\varphi = \cos(\varphi)\sigma_z + \sin(\varphi)\sigma_x$ for $\varphi \in (-\pi, \pi]$. Using the maximally entangled state $|\psi\rangle_{AB} = (|00\rangle + |11\rangle)/\sqrt{2}$ with the measurements depicted, one has a frequency distribution of $\boldsymbol{\omega}_{\mathcal{G}} = \frac{1}{2}\left(\frac{1}{2} + \frac{\sqrt{2}}{4}, 1, \frac{1}{2} - \frac{\sqrt{2}}{4}\right)$, where the scores are ordered $(c_{\text{CHSH}}, c_{\text{align}}, 0)$. The setup achieves Tsirelson's bound for the CHSH game as well as perfect correlations for the $X = 0$ and $Y = 2$ inputs. In addition, self-testing results [74] give a converse result: these scores completely characterize the devices up to local isometries. This implies that the state used by the devices is uncorrelated with an adversary and that the measurement pair $(X, Y) = (1, 2)$ yields uniformly random results, certifying the presence of 2 bits of private randomness in the outputs.

2.5.2 Loopholes

Loopholes in Bell-experiments are failures to exactly adhere to the given assumptions. In such a situation, the conclusions that we are able to draw from our observations are naturally subject to change. We shall now review two loopholes that will influence how we conduct the later analysis, for an in-depth discussion of the various loopholes in Bell-experiments we refer the reader to [75]. It should also be noted at this point that loophole-free[11] Bell-experiments have been performed [76–78].

Locality loophole

In a Bell-experiment we made the explicit assumption that, once separated, neither agent is capable of sending information to the other. This is often referred to as an assumption of *locality*: the result of one agent's experiment should not be influenced by the actions of another distant agent. This supersedes the assumption that the agents input

[11] Well, one can never rule out superdeterminism, but we won't discuss that here: the universe forbids it.

choices are independent of one another – this is also sometimes referred to separately as an assumption of *free choice*. The concept of a 'distant agent' can be made precise, we say agent B is *distant* from agent A if agent B is located at a spacetime point outside of the causal past of agent A.

When conducting a Bell-experiment we do not observe the underlying joint distribution that governs the observed statistics. Instead, we are only privy to an empirical distribution. Due to the effects of finite statistics, coupled with the fact that we cannot guarantee that the experiment's statistics were produced in an i.i.d. manner, we are unable to retroactively check that the locality assumption was upheld, e.g. by attempting to factor the distribution as per (2.5.1) and (2.5.2). Therefore, we must trust that the experiment performed upheld this assumption. In practice, this is done by forcing spacelike separation of the parties during interactions.[12]

Detection loophole

Suppose that during one of the trials of a Bell-experiment an input is provided to a device but the device fails to produce an output. What data should be recorded for this trial? Those of a non-conspiratorial disposition may be inclined to ignore the trial. However, such an action opens what is sometimes referred to as the *post-selection*, *detection* or *coincidence* loophole.[13] To highlight the problem, consider the following example wherein we have two agents play the CHSH game. Their strategy is simple, always output zero; except, when agent A receives the input 1 from the referee, then agent A refuses to respond. By ignoring these events, the only question pairs that contribute to the game score are $(0,0)$ and $(0,1)$ and for these question pairs a joint response of $(A,B)=(0,0)$ will result in a score of 1. By post-selecting on certain inputs, one can guarantee to win the game on all recorded runs.

In order to close the detection loophole one must take all of the trials into account. The most general method for doing so is to introduce a new symbol $\not{A}$ ($\not{B}$), which is recorded when Alice's (Bob's) detection event fails. This method increases the size of the output alphabets $\mathcal{A} \mapsto \mathcal{A} \cup \{\not{A}\}$ and $\mathcal{B} \mapsto \mathcal{B} \cup \{\not{B}\}$, opening up the possibility of using different Bell-inequalities. However, the original Bell-inequality may also be used by 'lifting' it to the new scenario [79]. Alternatively, instead of introducing a new symbol, one can record a pre-existing outcome upon observing a failed detection event. Whilst this is clearly less general than the previous method, there are contexts in which it may be advantageous to not increase the size of the output alphabets. For example, when optimizing over probability distributions achieving some Bell-inequality violation a larger alphabet results in a more

[12] In cryptographic scenarios we must assume that our devices cannot signal to an adversary, otherwise secrecy is non-existent. We can therefore extend this assumption to include that the devices within the laboratory do not signal to each other – if we can block the devices from signalling to an adversary then we should be able to block them from signalling to each other. As such, we do not look to close the locality loophole within our randomness expansion protocols.

[13] The various terms come from deficiencies in different implementations of Bell-experiments. However, they are all concerned with the same problem of incomplete data.

computationally intensive problem (see Appendix A).

2.6 Device-independence

Device-independence is a paradigm in which one attempts to draw conclusions about the outcomes of information processing tasks whilst relying as little as possible on the specific details of their implementation.[14] Instead, one looks to use readily available information to try and infer properties about the outcomes of some procedure. For example, as was noted in Remark 2.9, we may use the observation of nonlocal correlations to certify the existence of private randomness. Crucially, this can be done without reference to the actual system used to generate the nonlocal correlations, we need only to check that the assumptions required for a Bell-experiment are upheld to ensure the privacy.

Within a device-independent protocol, tasks are completed through a series of interactions with some *untrusted devices*. A *device* $\mathfrak{D}$ refers to some physical system that receives classical inputs and produces classical outputs. Describing such systems in the language of Def. 2.2, inputs correspond to a choice of state preparation and/or measurement procedure and the outputs would be the results of some measurements performed on the system. We say that $\mathfrak{D}$ is *untrusted* if the mechanism by which $\mathfrak{D}$ produces the outputs from the inputs is unknown. During the protocol, the agents interact with their untrusted devices within the following scenario:[15]

1. The protocol is performed within the confines of a secure lab from which information can be prevented from leaking.

2. This lab can be partitioned into two disconnected sites (one controlled by Alice and one by Bob).

3. The agents can send information freely and securely between these sites, Moreover, they are capable of preventing any unwanted information transfer between the sites.[16]

4. The agents each have their own device to which they can provide inputs (taken from alphabets $\mathcal{X}$ and $\mathcal{Y}$) and receive outputs (from alphabets $\mathcal{A}$ and $\mathcal{B}$).

5. These devices operate according to quantum theory, i.e., $\boldsymbol{p}_{AB|XY} \in \Omega_{\mathcal{AB}|\mathcal{XY}}$. Any eavesdropper is also limited by quantum theory. We use $\mathfrak{D}_{ABE}$ to denote the collection of devices (including any held by an eavesdropper).

[14] Of course, we must make some assumptions about the implementation. For example, in cryptographic schemes we assume that our secrets are not broadcast to an adversary, else cryptography is a moot concept. Rather, device-independence strives to make as few assumptions about the implementation as possible.

[15] One does not have to recreate this scenario exactly in order to perform the protocol. Instead, the given scenario establishes one situation in which the protocol remains secure.

[16] This imposes the locality restriction required for Bell-experiments.

6. The user has an initial source of private random numbers and a trusted computer for classical information processing.

One of the main advantages a device-independent protocol possesses over other cryptographic procedures, are its minimalist assumptions on the implementation. Moreover, through the observation of nonlocal correlations it has the capability of determining whether or not the untrusted devices are performing adequately. The protocols hence remain impervious to many side-channel attacks, malfunctioning devices or prior tampering.

2.6.1 The device-independent guessing probability problem

In the device-independent scenario we do not know the quantum states or measurements being performed. Instead, our entire knowledge about these must be inferred from the observed input-output behaviour of the devices used. In particular, observing correlations that violate a Bell-inequality provides a coarse-grained characterisation of the underlying system. As formulated in the definition of the min-entropy, the guessing probability (2.2.11) is not a device-independent quantity because its computation requires knowledge of the states ρ_E^x. However, the guessing probability can be reformulated in a device-independent way [37,38,80,81].

Consider a tripartite system ρ_{ABE} shared between two devices in the agents' lab and Eve. Because we are assuming an adversary limited by quantum theory, we can suppose that, upon receiving some inputs $(x,y) \in \mathcal{XY}$, the devices work by performing measurements $\{E_{a|x}\}_a$ and $\{F_{b|y}\}_b$ respectively, which give rise to some probability distribution $\boldsymbol{p} \in \mathfrak{Q}_{\mathcal{AB}|xy}$, and the overall state post-measurement is the cq-state

$$\sum_{ab} p(a,b|x,y)\,|a\rangle\langle a| \otimes |b\rangle\langle b| \otimes \rho_E^{abxy},$$

where

$$\rho_E^{abxy} = \frac{\mathrm{Tr}_{AB}\left[(E_{a|x} \otimes F_{b|y} \otimes \mathbb{1}_E)\rho_{ABE}\right]}{\mathrm{Tr}\left[(E_{a|x} \otimes F_{b|y} \otimes \mathbb{1}_E)\rho_{ABE}\right]}.$$

Note that the agents are not aware of the inner workings of the devices, they only observe the results of the measurements on their systems.

Consider the best strategy for Eve to guess the value of AB using her system E. She can perform a measurement on her system to try to distinguish the ensemble of states $\{\rho_E^{abxy}\}_{ab}$. Denoting Eve's POVM $\{M_e\}_e$ with outcomes in one-to-one correspondence with the values AB can take (say e_{ab} being the value corresponding to a best guess of $AB = (a,b)$)[17], then given some values of a,b,x and y, Eve's outcomes are distributed as $p(e_{a'b'}|a,b,x,y) = \mathrm{Tr}\left[M_{e_{a'b'}}\rho_E^{abxy}\right]$, and her probability of guessing correctly is

[17] Without loss of generality we can assume Eve's measurement has as many outcomes as what she is trying to guess.

$p(e_{ab}|a,b,x,y) = \mathrm{Tr}\left[M_{e_{ab}} \rho_E^{abxy}\right]$. For a fixed quantum strategy $q = \{\rho_{ABE}, \{E_{a|x}\}, \{F_{b|y}\}\}$, the overall probability of guessing AB correctly given E and $XY = (x,y)$ is

$$\begin{aligned} p_{\text{guess}}(AB|x,y,E,q) &= \sup_{\{M_e\}_e} \sum_{ab} \mathrm{Tr}\left[(E_{a|x} \otimes F_{b|y} \otimes M_{e_{ab}})\rho_{ABE}\right] \\ &= \sup_{\{M_e\}_e} \sum_{ab} p(a,b,e_{ab}|x,y,q) \\ &= \sup_{\{M_e\}_e} \sum_{ab} p(e_{ab}|a,b,x,y,q) p(a,b|x,y,q). \end{aligned}$$

Note that the guessing probability depends on the inputs x,y. In the protocols that we will consider later, there will only be one pair of inputs for which Eve is interested in guessing the outputs. We denote these inputs by $\tilde{x}$ and $\tilde{y}$.

In the device-independent scenario, Eve can also optimize over all quantum states and measurements that could be used by the devices. However, she wants to do so while restricting the devices to obey certain relations which the agents will be using to check the quality of their devices (for example, the agents may look for a CHSH violation greater than some value). For the moment, without specifying these relations precisely, call the set of quantum states and measurements obeying these relations $\mathcal{R}$. Hence, we seek

$$p_{\text{guess}}(AB|\tilde{x},\tilde{y},E) = \sup_{q\in\mathcal{R},\{M_e\}_e} \sum_{ab} p(a,b|\tilde{x},\tilde{y},q) p(e_{ab}|a,b,\tilde{x},\tilde{y},q).$$

Because Eve's measurement commutes with those of the devices (due to no signalling), we can use Bayes' rule to rewrite the optimization as[18]

$$\sup_{q\in\mathcal{R},\{M_e\}_e} \sum_{ab} p(e_{ab}|\tilde{x},\tilde{y},q) p(a,b|e_{ab},\tilde{x},\tilde{y},q).$$

With this rewriting it is evident that we can think about Eve's strategy as follows: Eve randomly chooses a value of $E = e$ and then prepares the device according to the choice e, trying to bias A,B towards the values a,b corresponding to the chosen e. We can hence write

$$p_{\text{guess}}(AB|\tilde{x},\tilde{y},E) = \sup_{\{\boldsymbol{p}_e\}_e} \sum_{ab} \mathbb{P}[E = e_{ab}]\, p_{e_{ab}}(a,b|\tilde{x},\tilde{y},q),$$

where $\sum_e p(e)\boldsymbol{p}_e$ satisfies some relations (equivalent to the restriction to the set $\mathcal{R}$) and $\boldsymbol{p}_e \in \mathfrak{Q}_{\mathcal{AB}|\mathcal{XY}}$ for each e.

The constraints imposed for the remainder of this thesis are those implied by assuming an expected frequency distribution $\boldsymbol{\omega}$ for some nonlocal game $\mathcal{G}$, e.g.

$$\sum_{abxy} \mu(x,y) p(a,b|x,y) \delta_{V(a,b,x,y),c} = \boldsymbol{\omega}(c)$$

for each $c \in \mathcal{G}$. Note that these constraints are linear functions of the probabilities $\boldsymbol{p}_e$. The optimization is then a conic program (the set of un-normalized quantum-realisable

[18] This rewriting makes sense provided no information leaks to Eve during the protocol, which is reasonable for randomness expansion since it takes place in a single secure laboratory.

distributions forms a convex cone). By introducing the subnormalised distributions $\tilde{\boldsymbol{p}}_e = \mathbb{P}[E=e]\boldsymbol{p}_e$, the problem can be expressed as

$$\begin{aligned} \sup_{\{\tilde{\boldsymbol{p}}_e\}_e} \quad & \sum_{ab} \tilde{p}_{e_{ab}}(a,b|\tilde{x},\tilde{y}) \\ \text{subj. to} \quad & \sum_e \sum_{abxy} \mu(x,y) p_e(a,b|x,y) \delta_{V(a,b,x,y),c} = \boldsymbol{\omega}(c) \qquad \forall c \in \mathcal{G} \\ & \tilde{\boldsymbol{p}}_e \in \tilde{\Omega}_{\mathcal{AB}|\mathcal{XY}} \quad \forall\, e\,. \end{aligned} \tag{2.6.1}$$

Note that the normalization constraint on $\sum_e \tilde{\boldsymbol{p}}_e$ is implied by the nonlocal game constraints as $\sum_c \boldsymbol{\omega}(c) = 1$.

Example 2.5. The guessing probability program for a pair of devices playing the CHSH game with an expected winning probability of $\omega \in \left[\frac{3}{4}, \frac{1}{2}+\frac{\sqrt{2}}{4}\right]$ is

$$\begin{aligned} \sup_{\{\tilde{\boldsymbol{p}}_e\}_e} \quad & \sum_{ab} \tilde{p}_{e_{ab}}(a,b|\tilde{x},\tilde{y}) \\ \text{subj. to} \quad & \frac{1}{4} \sum_{\substack{abxy: \\ a\oplus b = xy}} \sum_e p_e(a,b|x,y) = \omega \\ & \frac{1}{4} \sum_{\substack{abxy: \\ a\oplus b \neq xy}} \sum_e p_e(a,b|x,y) = 1-\omega \\ & \tilde{\boldsymbol{p}}_e \in \tilde{\Omega}_{\mathcal{AB}|\mathcal{XY}} \quad \forall\, e\,. \end{aligned}$$

Optimizing over the set of quantum correlations is a difficult problem, in part because the dimension of the quantum system achieving the optimum could be arbitrarily large. In order to get around this problem we consider a computationally tractable relaxation of the problem, which the next subsection is dedicated to introducing.

2.7 Semidefinite programming

This final section of the preliminaries covers the topic of semidefinite programs (SDPs). We will briefly introduce the general topic before covering a particular application which will allow us to make the optimization problem (2.6.1) computationally tractable. We denote the set of symmetric $n \times n$ real-valued matrices by $\mathfrak{S}_n(\mathbb{R})$.

2.7.1 The basics

Definition 2.9 (Primal SDP)**.** Let $C, F_1, \ldots, F_r \in \mathfrak{S}_n(\mathbb{R})$ and $b_1, \ldots, b_r \in \mathbb{R}$. The collection $(C, F_1, \ldots, F_r, b_1, \ldots, b_r)$ defines the optimization problem

$$\begin{aligned} \sup_{X \in \mathfrak{S}_n(\mathbb{R})} \quad & \mathrm{Tr}[CX], \\ \text{subject to} \quad & \mathrm{Tr}[F_i X] = b_i \quad \text{for all } i \in 1, \ldots, r, \\ & X \geq 0, \end{aligned} \tag{2.7.1}$$

which we refer to as a *semidefinite program* (SDP) in its *primal* form.

By the principle of Lagrangian duality [82] there exists a secondary optimization problem, known as the *dual problem*, which offers an alternate perspective on the optimization. It may be derived by considering the Lagrangian of (2.7.1)

$$L(X,\boldsymbol{\lambda},Y) = \mathrm{Tr}[CX] + \sum_i \lambda_i (b_i - \mathrm{Tr}[F_i X]) + \mathrm{Tr}[XY] \tag{2.7.2}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^r$ and $Y \succeq 0$ are *dual variables*. The primal problem is recovered from the Lagrangian by considering the *primal functional* $f(X) := \inf_{\boldsymbol{\lambda},Y} L(X,\boldsymbol{\lambda},Y)$. The constraints imposed by the primal problem emerge from requiring that $f(X)$ is bounded. That is, for $X \in \mathfrak{S}_n(\mathbb{R})$ we have

$$\inf_{Y\succeq 0} \mathrm{Tr}[XY] = \begin{cases} 0 & \text{if } X \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \tag{2.7.3}$$

and

$$\inf_{\lambda_i \in \mathbb{R}} \lambda_i (b_i - \mathrm{Tr}[F_i X]) = \begin{cases} 0 & \text{if } \mathrm{Tr}[F_i X] = b_i \\ -\infty & \text{otherwise.} \end{cases} \tag{2.7.4}$$

Therefore, by enforcing that $\sup_{X\in\mathfrak{S}_n(\mathbb{R})} f(X)$ is bounded from below we recover precisely the primal problem.

We derive the dual problem by switching the order with which we take the supremum and infimum of the Lagrangian. By the linearity of the trace, we may rearrange the Lagrangian

$$L(X,\boldsymbol{\lambda},Y) = \sum_i \lambda_i b_i + \mathrm{Tr}\left[(C - \sum_i \lambda_i F_i + Y)X \right],$$

and we define the *dual functional* $g(\boldsymbol{\lambda},Y) := \sup_X L(X,\boldsymbol{\lambda},Y)$. As before, the constraints of the problem emerge from the requirement that the functional is bounded. This is precisely when $C - \sum_i \lambda_i F_i + Y = 0$ and as we are already imposing $Y \succeq 0$, this is equivalent to the condition $C - \sum_i \lambda_i F_i \preceq 0$ with $Y = \sum_i \lambda_i F_i - C$ now implicit. Subject to these conditions holding, the Lagrangian dual functional becomes $g(\boldsymbol{\lambda}) = \sum_i \lambda_i b_i$ and we arrive at the dual formulation of our problem.

Definition 2.10 (Dual SDP)**.** Let $C, F_1, \ldots, F_r \in \mathfrak{S}_n(\mathbb{R})$ and $b_1, \ldots, b_r \in \mathbb{R}$. The collection $(C, F_1, \ldots, F_r, b_1, \ldots, b_r)$ defines the optimization problem

$$\begin{aligned} \inf_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad & \boldsymbol{\lambda}\cdot\boldsymbol{b} \\ \text{subject to} \quad & C - \sum_i \lambda_i F_i \preceq 0 \end{aligned} \tag{2.7.5}$$

which we refer to as a *semidefinite program* (SDP) in its *dual* form.

Remark 2.11. We refer to a matrix $X' \in \mathfrak{S}_n(\mathbb{R})$ or a vector $\boldsymbol{\lambda}' \in \mathbb{R}^m$ as *feasible points* of the primal and dual SDPs respectively, if they satisfy the constraints imposed by the respective optimization problems. Additionally, any such points that satisfy the respective strict versions of the inequality constraints are referred to as *strictly feasible points*. If the set of feasible points of an optimization problem is empty, then we refer to that problem as *infeasible* and consider its value to be $-\infty$ if the problem's objective is a supremum and ∞ if it is an infimum.

We shall now review several properties of the primal and dual formulation of SDPs.

- **Weak duality**: SDPs (and all optimization problems in general) satisfy a property known as *weak duality*. That is, if p^* and d^* denote the optimal values of the primal and dual problems, respectively. Then we have the ordering

$$p^* \leq d^*. \tag{2.7.6}$$

  This follows immediately from the construction of the primal and dual programs from the Lagrangian, together with the *max-min inequality* [82]: for any $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and any $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^m$ we have

$$\sup_{\boldsymbol{v}\in V} \inf_{\boldsymbol{w}\in W} f(\boldsymbol{v},\boldsymbol{w}) \leq \inf_{\boldsymbol{w}\in W} \sup_{\boldsymbol{v}\in V} f(\boldsymbol{v},\boldsymbol{w}). \tag{2.7.7}$$

- **Strong duality**: We say that the primal and dual formulations of the optimization problem are *strongly dual*, whenever $p^* = d^*$. Unlike linear programming, strong duality may not always occur in feasible SDPs. However, there exist sufficient conditions for strong duality to hold (more generally known as constraint qualification conditions). One such condition is *Slater's condition* [83], which states that if one of the formulations is strictly feasible then the other achieves its optimum exactly and the two formulations as a pair are strongly dual.

- **Inequality constraints**: Note that the primal and dual formulations do not contain any linear inequality constraints. However, any inequality constraint can be implemented using an equality constraint together with an augmentation of the semidefinite constraint. That is, to implement the constraint $\mathrm{Tr}[FX] \leq b$, we can introduce an additional variable $s \in \mathbb{R}$ into the problem, known as a *slack variable*. We then rewrite the inequality as the equality $\mathrm{Tr}[FX] = b - s$ together with $s \geq 0$. The constraint $s \geq 0$ can be implemented by modifying the semidefinite constraint $X \succeq 0$ to be $X \bigoplus s \succeq 0$.[19]

[19] Depending on the algorithm used to solve the SDP, all equality constraints may be converted to two inequality constraints, i.e. $\mathrm{Tr}[FX] = b$ becomes $\mathrm{Tr}[FX] \leq b$ and $\mathrm{Tr}[FX] \geq b$. The constraints imposed by the program can then be viewed as one large semidefinite constraint. This conversion is used in the primal-dual interior point method implemented by the sdpa solvers [84].

- **Dual functional as a bounding hyperplane**: Let $D(\boldsymbol{b}) := \{X \in \mathfrak{S}_n(\mathbb{R}) \mid \mathrm{Tr}[F_i X] = b_i, X \succeq 0\}$ be the set of feasible points for the primal problem (2.7.1) parametrized by the vector $\boldsymbol{b} \in \mathbb{R}^r$, let $p^*(\boldsymbol{b}) = \sup_{X \in D(\boldsymbol{b})} \mathrm{Tr}[CX]$ denote the optimum primal value and let $\boldsymbol{\lambda_b}$ be a feasible point in the dual program. Then define the function $g_{\boldsymbol{b}} : \boldsymbol{c} \mapsto \boldsymbol{\lambda_b} \cdot \boldsymbol{c}$ for any $\boldsymbol{c} \in \mathbb{R}^r$. Then, for any $\boldsymbol{c} \in \mathbb{R}^r$ we have

$$p^*(\boldsymbol{c}) \leq g_{\boldsymbol{b}}(\boldsymbol{c}). \tag{2.7.8}$$

  That is, a dual functional derived from any parametrization of the dual program[20] can be evaluated to give an upper bound on the optimal solution to a primal problem. To see this, take the dual constraint for the program parametrized by $\boldsymbol{b}$, i.e. $C - \sum_i \lambda_i F_i \preceq 0$, and for some $X \in D(\boldsymbol{c})$ apply the map $M \mapsto \mathrm{Tr}[MX]$ for $M \in \mathfrak{S}_n(\mathbb{R})$. This implies,

$$\mathrm{Tr}[CX] \leq \sum_i \lambda_i c_i$$

  and (2.7.8) follows from taking the supremum over all $X \in D(\boldsymbol{c})$.

2.7.2 Semidefinite relaxations of quantum correlations

Given the ubiquity of positive semidefinite matrices within the field quantum information, it comes with little to no surprise to find that SDPs have wide ranging applications in the area [46]. In this thesis, we are interested in how SDPs can be used to approximate the set of quantum correlations $\mathfrak{Q}_{\mathcal{AB}|\mathcal{XY}}$. This allows us to compute quantities such as the device-independent guessing probability (2.6.1) without having to directly optimise over $\mathfrak{Q}$. Interestingly, this particular application of semidefinite programming did not begin with quantum information in mind. Rather, it started with the problem of optimizing multivariate polynomials [85, 86] and the techniques developed were later generalised to the case of non-commutative multivariate polynomials [87]. The latter problem can then be used to approximate the set of quantum correlations [40, 41].

Recall that for a distribution $\boldsymbol{p} \in \mathfrak{P}_{\mathcal{AB}|\mathcal{XY}}$ to be *quantum*, there must exist a Hilbert space $\mathcal{H}$, a state $\rho \in \mathcal{S}(\mathcal{H})$ and measurement operators $\{\{E_{a|x}\}_{a \in \mathcal{A}}\}_{x \in \mathcal{X}}, \{\{F_{b|y}\}_{b \in \mathcal{B}}\}_{y \in \mathcal{Y}} \in \mathcal{M}(\mathcal{H})$ such that $p(a,b|x,y) = \mathrm{Tr}\left[\rho E_{a|x} \otimes F_{b|y}\right]$ for all $(a,b,x,y) \in \mathcal{AB}\mathcal{XY}$. The purpose of this section is to define a collection of converging outer approximations $\mathfrak{Q}^{(1)} \supseteq \mathfrak{Q}^{(2)} \supseteq \cdots \supseteq \mathfrak{Q}$, such that membership of $\mathfrak{Q}^{(k)}$ is equivalent to the existence of some positive semidefinite matrix $\Gamma^{(k)}$, this hierarchy of necessary conditions is sometimes referred to as the NPA hierarchy. Therefore, given an optimization problem we can replace membership of $\mathfrak{Q}$ with the computationally more appealing constraint, the existence of a positive semidefinite matrix, to get a bound on the solution of the problem.

[20] The dual functional will however be trivial if the parametrization has an empty feasible set.

As we place no dimension restriction on $\mathcal{H}$,[21] the Naimark and Stinespring dilation theorems[22] [88] tell us that without loss of generality we may restrict our considerations to that of pure states and projective measurements. Let $\mathcal{W}^{(1)} = \{\mathbb{1}\} \cup \{E_{a|x}\}_{(a,x)\in\mathcal{A}\mathcal{X}} \cup \{F_{b|y}\}_{(b,y)\in\mathcal{B}\mathcal{Y}}$. We refer to a product of elements of $\mathcal{W}^{(1)}$ as a *word*, e.g. $w = E_{a|x}F_{b|y}E_{a'|x}$ is a word. Due to algebraic constraints on the operators; e.g. impotency $E_{a|x}^2 = E_{a|x}$, commutativity $E_{a|x}F_{b|y} = F_{b|y}E_{a|x}$ and orthogonality $E_{a|x}E_{a'|x} = 0$, many words will be equivalent. Formally, we can define an equivalence relation $\sim$ where $w_1 \sim w_2$ if they correspond to the same operator. We define the *length* of a word w, denoted $|w|$, to be the shortest product of elements in the equivalence class $[w]$. For example, $|\mathbb{1}E_{a|x}F_{b|y}E_{a|x}F_{b|y}| = |E_{a|x}F_{b|y}| = 2$.

Now we define the sets $\mathcal{W}^{(k)}$ for $k \in \mathbb{N}$ to be all words, up to equivalences, of length no greater that k. For example, one possible identification of $\mathcal{W}^{(2)}$ is

$$\mathcal{W}^{(2)} = \mathcal{W}^{(1)} \cup \{E_{a|x}E_{a'|x'}\}_{a,a'\in\mathcal{A},x,x'\in\mathcal{X}:x\neq x'} \cup \{F_{b|y}F_{b'|y'}\}_{b,b'\in\mathcal{B},y,y'\in\mathcal{Y}:y\neq y'} \cup \{E_{a|x}F_{b|y}\}_{(a,b,x,y)\in\mathcal{A}\mathcal{B}\mathcal{X}\mathcal{Y}}.$$

We refer to $\mathcal{W}^{(k)}$ as the *monomial set of level* k. Now given $\mathcal{W}^{(k)}$ we define the *certificate of level* k to be the matrix $\Gamma^{(k)}$ indexed by elements of $\mathcal{W}^{(k)}$ such that

$$\Gamma^{(k)}_{v,w} = \mathrm{Tr}\left[\rho v^\dagger w\right], \tag{2.7.9}$$

for each $v, w \in \mathcal{W}^{(k)}$ and some $\rho \in \mathcal{S}(\mathcal{H})$. Then we say a distribution $\boldsymbol{p} \in \mathfrak{P}$ has a certificate of level k, denoted by $\boldsymbol{p} \in \mathfrak{Q}^{(k)}$, if there exists a $\Gamma^{(k)}$ with entries that are consistent with $\boldsymbol{p}$, i.e. the following all hold

$$\Gamma^{(k)}_{\mathbb{1},E_{a|x}} = p(a|x), \tag{2.7.10}$$

$$\Gamma^{(k)}_{\mathbb{1},F_{b|y}} = p(b|y), \tag{2.7.11}$$

$$\Gamma^{(k)}_{E_{a|x},F_{b|y}} = p(a,b|x,y). \tag{2.7.12}$$

Remark 2.12. As each level of the hierarchy adds additional constraints to the problem we have the inclusion chain $\mathfrak{Q}^{(1)} \supseteq \mathfrak{Q}^{(2)} \supseteq \ldots$. Moreover, the hierarchy provides necessary and sufficient conditions for a behaviour to be quantum: $\boldsymbol{p} \in \mathfrak{Q}$ iff $\boldsymbol{p} \in \mathfrak{Q}^{(k)}$ for all $k \in \mathbb{N}$.

The conic program (2.6.1) now becomes a semidefinite program which can be solved in an efficient manner, at the expense of possibly not obtaining the same optimum value. The corresponding relaxed problem is

$$\begin{aligned} p^{(k)}_{\text{guess}}(\boldsymbol{\omega}) := \sup_{\{\tilde{\boldsymbol{p}}_e\}_e} \quad & \sum_{ab} \tilde{p}_{e_{ab}}(a,b|\tilde{x},\tilde{y}) \\ \text{subj. to} \quad & \sum_e \sum_{abxy} \mu(x,y) p_e(a,b|x,y) \delta_{V(a,b,x,y),c} = \omega(c) \qquad \forall\, c \in \mathcal{G} \\ & \tilde{\boldsymbol{p}}_e \in \tilde{\mathfrak{Q}}^{(k)} \quad \forall\ e\,. \end{aligned} \tag{2.7.13}$$

[21] The NPA hierarchy also applies to infinite dimensional Hilbert spaces.

[22] Both dilation theorems ascertain the existence of an isometry $V : \mathcal{H} \to \mathcal{H}'$, 'dilating the Hilbert space', such that the objects of interest under this mapping take an arguably simpler form. In particular, Naimark's theorem dilates POVMs to projective measurements and Stinespring's theorem dilates completely positive maps to unitary maps.

where we have now explicitly parametrized the program by some expected frequency distribution $\boldsymbol{\omega} \in \mathfrak{Q}_G$. Since the NPA hierarchy forms a sequence of outer approximations to the set of quantum correlations, $\mathfrak{Q}_1 \supseteq \mathfrak{Q}_2 \supseteq \cdots \supseteq \mathfrak{Q}$, the relaxed guessing probability provides an upper bound on the true guessing probability, i.e., $p_{\text{guess}}(\boldsymbol{\omega}) \leq p^{(k)}_{\text{guess}}(\boldsymbol{\omega})$. Combined with (2.2.12), one can use the relaxed programs to compute valid device-independent lower bounds on $H_{\min}$. We denote a feasible point of the dual of (2.7.13), when parametrized by $\boldsymbol{\omega}$, by $\boldsymbol{\lambda}_{\boldsymbol{\omega}}$. Note that for our later analysis we only need $\boldsymbol{\lambda}_{\boldsymbol{\omega}}$ to be a feasible point of the dual program, we do not require it to be optimal.[23] Any feasible point allows us to construct functions that upper bound the guessing probability (cf. (2.7.8)) and in turn lower bound the min-entropy.

[23]The optimal point may not even be achievable unless the program is strongly dual.

CHAPTER 3

A FRAMEWORK FOR CONSTRUCTING RANDOMNESS EXPANSION PROTOCOLS

This chapter presents a method for constructing quantum-secure randomness expansion protocols. We begin by introducing the task of randomness expansion and the generic spot-checking protocol. In Sec. 3.2 we present a numerical construction of min-tradeoff functions and show how this can be applied to the spot-checking protocol to generate the relevant security statements. We conclude the chapter in Sec. 3.3 with an analysis of the cost of seeding the protocol.

3.1 Randomness expansion

A device-independent randomness expansion protocol is a procedure by which one attempts to use a private and uniform seed to produce a longer private and uniform output, through repeated interactions with some untrusted devices. We begin by introducing the generic structure of the *spot-checking* protocol which we will build upon to produce our quantum randomness expansion (QRE) protocol (see Fig. 3.3).

3.1.1 The spot-checking protocol

Randomness expansion consists of three main subprocedures: accumulation, evaluation and extraction. During the accumulation phase, agents interact with their untrusted devices in an attempt to generate entropy. Next, the evaluation phase acts as the quality control step in the protocol. During evaluation the statistics produced during the accumulation step are checked against some preselected nonlocality test. If the untrusted devices fail to display a sufficient level of nonlocality then the protocol is abandoned. However, if the protocol does not abort then the agents can use the EAT to place a probabilistic lower bound on the

private randomness produced during the accumulation phase. Finally, with a lower bound on the total entropy produced the agents can apply a randomness extractor to compress the long, partially random output string into a shorter string of almost uniformly random bits. Let us now elaborate on these three subprocedures.

Accumulation

During the accumulation step, the agents interact with their respective devices in order to generate randomness. Before beginning this subprocedure the agents place their untrusted devices in secure laboratories subject to the conditions detailed in Sec. 2.6. Then, a nonlocality test $\mathcal{G} = (\mu, V)$ is chosen, which will be used to evaluate the quality of the devices in the next step.

The accumulation procedure consists of $n \in \mathbb{N}$ separate interactions with the untrusted devices. We refer to a single interaction with the devices as a *round*. A round consists of the agents selecting and supplying inputs to the devices, receiving outputs and recording this data. During the i^{th} round, a random variable $T_i \sim \text{Bernoulli}(\gamma)$ is sampled, for some fixed $\gamma \in (0,1)$, indicating whether the round will be a *generation round* or a *test round*. With probability $1-\gamma$ we have $T_i = 0$ and the round is a generation round. During a generation round, the agents supply their respective devices with the fixed generation inputs $(\tilde{x}, \tilde{y}) \in \mathcal{X}\mathcal{Y}$, recording $X_i Y_i = (\tilde{x}, \tilde{y})$. They record the outputs they receive from their devices as A_i and B_i respectively and they record the round's score as $C_i = \perp$. With probability γ, $T_i = 1$ and the round is a test round. During a test round, inputs $X_i Y_i$ are sampled according to the distribution specified by the nonlocal game. The sampled inputs are fed to their respective devices and the outputs received are recorded as $A_i B_i$. The score is computed and recorded as $C_i = V(A_i, B_i, X_i, Y_i)$. The *transcript for round* i is the tuple $(A_i, B_i, X_i, Y_i, T_i, C_i)$. After n rounds, the complete transcript for the accumulation procedure is $(A_1^n, B_1^n, X_1^n, Y_1^n, T_1^n, C_1^n)$.

Remark 3.1. Let us expand on several aspects of the accumulation procedure.

1. The generation inputs should be chosen in order to maximize the randomness generated, e.g. in the extended CHSH game (cf. Fig. 2.2) one should choose $(\tilde{x}, \tilde{y}) = (1, 2)$.

2. By choosing the rounds randomly according to a distribution heavily favouring generation rounds, we are able to reduce the size of the seed whilst sufficiently constraining the device's behaviour, guaranteeing the presence of randomness within the outcomes (except with some small probability). This allows us to maximize our net gain in entropy.

3. From the description of the accumulation procedure above, it may see that the spot checking protocol is not compatible with a loophole free test of nonlocality. As information about the round type propagates to the agents, which influences their

choice of input, it is entirely possible that this information could also propagate to the devices and influence their behaviour. This would render moot any inference about the devices' behaviour during generation rounds from observations of its behaviour during test rounds. This can be remedied by either introducing the additional assumption that the devices are shielded from this information (see Sec. 2.6) or by selecting the round types in a manner that avoids this loophole (see the supplementary material of [89]).

Evaluation

The accumulation step sees the agents produce a transcript $(A_1^n, B_1^n, X_1^n, Y_1^n, T_1^n, C_1^n)$. Next, they look to determine the quality of this transcript and, in turn, certify a lower bound on the total entropy produced, $H_{\min}^{\epsilon}(A_1^n B_1^n | X_1^n Y_1^n E)$. To this end, the agents compute their *empirical frequency distribution*

$$F_{C_1^n}(c) = \frac{1}{n} \sum_{i=1}^{n} \delta_{c,C_i}. \tag{3.1.1}$$

Prior to the accumulation step, the agents fix some frequency distribution $\boldsymbol{\omega}$ that they expect (or hope) the devices to behave like in each round. Should the devices actually behave in an i.i.d. manner according to $\boldsymbol{\omega}$, then concentration bounds tell us that the empirical frequency distribution $\boldsymbol{F}_{C_1^n}$ should be close to this. With this in mind, we define the event that the protocol does not abort by

$$\Omega = \{C_1^n \mid \gamma(\boldsymbol{\omega}(\mathcal{G}) - \boldsymbol{\delta}) < \boldsymbol{F}_{C_1^n}(\mathcal{G}) < \gamma(\boldsymbol{\omega}(\mathcal{G}) + \boldsymbol{\delta})\}, \tag{3.1.2}$$

where $\boldsymbol{\delta} \in (0,1)^{|\mathcal{G}|}$ is a vector of confidence interval widths satisfying $\mathbf{0} < \boldsymbol{\delta} < \boldsymbol{\omega}(\mathcal{G})$ with all vector inequalities being interpreted as element-wise constraints.

If the agents have not aborted the protocol, then they may apply the EAT conditioned on the event Ω to lower bound the total smooth min-entropy produced during the accumulation phase. In order to do this the agents require a min-tradeoff function for their accumulation procedure. A method by which one can construct these min-tradeoff functions is the focus of Sec. 3.2.1.

Remark 3.2.

1. In order to fix a sensible expected frequency distribution prior to accumulation, the agents must have some knowledge about the expected behaviour of the devices. In practice, this could be communicated by the manufacturer of the devices. Note that a malicious manufacturer does not gain anything by providing the agents with an inaccurate behaviour. Such an action would only lead to a larger probability of abort.[1]

[1] And, in turn, one would expect a large reduction in sales.

2. The success event Ω does not constrain the value of $F_{C_1^n}(\perp)$. This is because the value of $F_{C_1^n}(\perp)$ is not indicative of the devices' nonlocal behaviour. Moreover, as the sampling of the test rounds is a trusted i.i.d. procedure the value $F_{C_1^n}(\perp)$ takes should, except with some small probability, be within a small region centred about $(1-\gamma)$.

3. The success event Ω consists of upper and lower bounds on each element of the frequency distribution. If the certifiable entropy is monotonic with respect to an element of the frequency distribution then we only require a one sided bound.[2] For example, a higher score in the CHSH game allows us to certify more entropy and so it is sufficient to assume just a lower bound (see (4.3.11)).

Extraction

If the protocol does not abort during the evaluation stage, then the agents will have produced a string of bits $A_1^n B_1^n$ that they are confident has at least k bits of smooth min-entropy, i.e. except with some small probability we have $H_{\min}^{\epsilon_s}(A_1^n B_1^n | X_1^n Y_1^n E) > k$. They are now free to apply a quantum-secure randomness extractor (see Sec. 2.4) to $A_1^n B_1^n$ to produce approximately k, close to uniformly random bits.

Remark 3.3. There is a question of whether the quantity we are actually interested in is $H_{\min}^{\epsilon_s}(A_1^n B_1^n | X_1^n Y_1^n E)$, rather than $H_{\min}^{\epsilon_s}(A_1^n B_1^n | E)$ or $H_{\min}^{\epsilon_s}(A_1^n B_1^n X_1^n Y_1^n | E)$. In common key-distribution protocols (such as BB84), the first of these is the only reasonable choice because the information $X_1^n Y_1^n$ is communicated between the two parties over an insecure channel and hence could become known by Eve. For randomness expansion, this is no longer the case: this communication can all be kept secure within a single laboratory. Whether the alternative quantities can be used then depends on where the seed randomness comes from. If a trusted beacon is used then the first case is needed. If the seed randomness can be kept secure until such time that the random numbers need no longer be kept random then the second quantity could be used[3]. If it is also desirable to extract as much randomness as possible, then the third quantity could be used instead. However, due to the spot-checking structure of the protocols, the amount of seed required for the entropy accumulation procedure is small enough that its reuse will not be of practical significance (see the discussion in Sec. 3.3).

[2] In fact, as min-tradeoff functions are required to be affine, they will necessarily be monotonic in each element of the empirical frequency distribution. Therefore, after selecting a min-tradeoff function one could modify Ω to only include one sided bounds. In turn, one could gain approximately a factor of two in the completeness error. In spite of this we present Ω in manner seen above as it is conceptually clearer and requires no additional knowledge of the min-tradeoff function structure.

[3] This is a reasonable requirement, because there are other strings that have to be kept secure in the same way, e.g., the raw string A_1^n.

3.1.2 Security definitions

The security definitions for our randomness expansion protocols are based on the distinguishability paradigm introduced in [90, 91] for classical cryptography. The idea is that one can define an ideal system that performs our cryptographic task securely. Then, we say a real system is secure if it is, except with some small probability, indistinguishable from the ideal system. This notion of security allows for the composition of secure protocols, without compromising the security of the overall procedure. To make this more precise, consider a hypothetical device that outputs a string Z that is uniform and uncorrelated with any information held by an eavesdropper. In other words, it outputs $\tau_m \otimes \rho_E$, where τ_m is the maximally mixed state on m qubits. The ideal protocol is defined as the protocol that involves first doing the real protocol, then, in the case of no abort, replacing the output with a string of the same length taken from the hypothetical device. The protocol is then said to be $\varepsilon_{\text{sound}}$-secure if, when the user either implements the real or ideal protocol with probability $\frac{1}{2}$, the maximum probability that a distinguisher can guess which is being implemented is at most $\frac{1+\varepsilon_{\text{sound}}}{2}$. If $\varepsilon_{\text{sound}}$ is small, then the real and ideal protocols are virtually indistinguishable. For an overview of composable security, with a focus on quantum cryptography, we refer the reader to [92].

The soundness error alone does not capture all of the features of a secure protocol. For example, one can construct a vacuously secure randomness expansion protocol by demanding that during the evaluation step the protocol always aborts. In this case, the real and ideal protocols are indistinguishable as they will both only ever abort. However, this is clearly not a very useful procedure. To avoid these scenarios we have a second security parameter, the *completeness error*, which is the probability that an ideal implementation of the protocol leads to an abort. Combining soundness and completeness we arrive at the security definition for our protocol.

Definition 3.1. Let $\mathcal{R}$ be a randomness expansion protocol producing an output Z and let Ω be the event that the protocol does not abort. Then, we say $\mathcal{R}$ is an $(\varepsilon_{\text{sound}}, \varepsilon_{\text{comp}})$-randomness expansion protocol if it satisfies the following two conditions.

1. **Soundness**:
$$\frac{1}{2}\mathbb{P}[\Omega] \cdot \|\rho_{ZE} - \tau_m \otimes \rho_E\|_1 \leq \varepsilon_{\text{sound}}, \tag{3.1.3}$$
where E is an arbitrary quantum register (which could have been entangled with the devices used at the start of the protocol), m is the length of the output string Z and τ_m is the maximally mixed state on a system of dimension 2^m.

2. **Completeness**: There exists a set of quantum states and measurements such that if they are used to implement protocol $\mathcal{R}$ then
$$\mathbb{P}[\Omega] \geq 1 - \varepsilon_{\text{comp}}. \tag{3.1.4}$$

Remark 3.4. Although we use a composable security definition to ensure that any randomness output by the protocol can be used in any scenario, importantly, this may not apply if the devices used in the protocol are subsequently reused [93]. Thus, after the protocol the devices should be kept shielded and not reused until the randomness generated no longer needs to be kept secure. How to best resolve this remains an open problem: the Supplemental Material of [93] presents ideas for modifications to the protocol (and modifications to the notion of composability) that may circumvent such problems.

3.2 A template randomness expansion protocol

In this section we introduce a method for constructing randomness expansion protocols that can be tailored to devices with different specifications. Our template protocol, Protocol QRE (see Fig. 3.3), follows the general spot-checking structure introduced above and is proven to be secure when used with any nonlocal game $\mathcal{G}$. We begin by showing how min-tradeoff functions can be constructed numerically and then apply this to the task of randomness expansion.

3.2.1 Numerical constructions of min-tradeoff functions

We now present a constructible family of min-tradeoff functions for a general instance of Protocol QRE. The construction is based on the following idea. As noted in Sec. 2.6.1 one can numerically calculate a lower bound on the min-entropy of a system based on its observed statistics. Pairing this with the relation, $H_{\min}(X|E) \leq H(X|E)$, we have access to numerical bounds on the von Neumann entropy. In particular, we can extract a linear functional from the dual of program (2.7.13), in order to construct a min-tradeoff function for the protocol.[4] However, in order to capture the spot-checking structure of the protocol, we must extend the domain of our function to include the no-test symbol $\perp$. We perform this extension by following the procedure detailed in [25].

As the rounds are split into testing and generation rounds, we may decompose the EAT-channel for the i^{th} round as $\mathcal{N}_i = \gamma\mathcal{N}_i^{\text{test}} + (1-\gamma)\mathcal{N}_i^{\text{gen}}$, where $\mathcal{N}_i^{\text{test}}$ is the channel that is applied if the round is a test round and $\mathcal{N}_i^{\text{gen}}$ if the round is a generation round. Importantly, this splitting separates the no-test symbol $\perp$ from the nonlocal game scores. That is, if $\mathcal{N}_i^{\text{test}}$ is applied then $\mathbb{P}[C_i = \perp] = 0$ whereas if $\mathcal{N}_i^{\text{gen}}$ is applied then $\mathbb{P}[C_i = \perp] = 1$. The following lemma, Lemma 5.5 in [25], explains how one can extend the domain of our entropy bounding functions to capture the spot-checking structure of our protocols.

[4] In fact, by relaxing the dual program to some level of the NPA hierarchy, the single round bound is valid against super-quantum adversaries. However, the security of the full protocol may not extend to such adversaries: to show that we would need to generalise the EAT and the extractor.

Lemma 3.1 (Min-tradeoff extension [25])**.** *Let $g : \mathfrak{P}_{\mathcal{G}} \to \mathbb{R}$ be an affine function satisfying*

$$g(\boldsymbol{p}) \leq \inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i^{test}(\sigma)_{C_i(\mathcal{G})} = \tau_{\boldsymbol{p}}} H(A_iB_i|X_iY_iR')_{\mathcal{N}_i(\sigma)} \tag{3.2.1}$$

for all $\boldsymbol{p} \in \mathfrak{Q}_{\mathcal{G}}$. Then, the function $f : \mathfrak{P}_{\mathcal{G}\cup\{\perp\}} \to \mathbb{R}$, defined by its action on trivial distributions

$$\begin{aligned} f(\boldsymbol{\delta}_c) &= \mathrm{Max}[g] + \frac{g(\boldsymbol{\delta}_c) - \mathrm{Max}[g]}{\gamma}, \qquad \forall c \in \mathcal{G}, \\ f(\boldsymbol{\delta}_\perp) &= \mathrm{Max}[g], \end{aligned}$$

is a min-tradeoff function for the EAT-channels $\{\mathcal{N}_i\}_i$. Furthermore, f satisfies the following properties:

$$\begin{aligned} \mathrm{Max}[f] &= \mathrm{Max}[g], \\ \mathrm{Min}_{\Sigma}[f] &\geq \mathrm{Min}[g], \\ \mathrm{Var}_{\Sigma}[f] &\leq \frac{(\mathrm{Max}[g] - \mathrm{Min}[g])^2}{\gamma}. \end{aligned}$$

We now have all the relevant machinery to state our numerical construction of min-tradeoff functions. The following lemma details precisely how one can use the relaxed dual of the guessing probability program in order to construct a min-tradeoff function for Protocol QRE.

Lemma 3.2 (Min-tradeoff construction)**.** *Let $\mathcal{G}$ be a nonlocal game and $k \in \mathbb{N}$. For each $\boldsymbol{\nu} \in \mathfrak{Q}_{\mathcal{G}}^{(k)}$, let $\boldsymbol{\lambda}_{\boldsymbol{\nu}}$ be a feasible point of the dual of Prog. (2.7.13) when parameterized by $\boldsymbol{\nu}$ and computed at the k^{th} relaxation level. Furthermore, let $\lambda_{\max} = \max_{c\in\mathcal{G}} \lambda_{\boldsymbol{\nu}}(c)$ and $\lambda_{\min} = \min_{c\in\mathcal{G}} \lambda_{\boldsymbol{\nu}}(c)$. Then, for any set of EAT channels $\{\mathcal{N}_i\}_{i=1}^n$ implementing an instance of Protocol QRE with the nonlocal game $\mathcal{G}$, the set of functionals $\mathcal{F}_{\min}(\mathcal{G}) = \left\{ f_{\boldsymbol{\nu}}(\cdot) \,|\, \boldsymbol{\nu} \in \mathfrak{Q}_{\mathcal{G}}^{(k)} \right\}$ forms a family of min-tradeoff functions, where $f_{\boldsymbol{\nu}} : \mathfrak{P}_{\mathcal{C}} \to \mathbb{R}$ are defined by*

$$f_{\boldsymbol{\nu}}(\boldsymbol{\delta}_c) := (1-\gamma)\left(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \frac{\lambda_{\boldsymbol{\nu}}(c) - (1-\gamma)\lambda_{\min}}{\gamma} \right) \quad \text{for } c \in \mathcal{G} \tag{3.2.2}$$

and

$$f_{\boldsymbol{\nu}}(\perp) := (1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\, \lambda_{\min}), \tag{3.2.3}$$

where $A_{\boldsymbol{\nu}} = \frac{1}{\ln 2} - \log(\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu})$ and $B_{\boldsymbol{\nu}} = \frac{1}{\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu} \ln 2}$ are constants defined by the solution to the dual program.

Moreover, these min-tradeoff functions satisfy the following relations.

- *Maximum:*

$$\mathrm{Max}[f_{\boldsymbol{\nu}}] = (1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\, \lambda_{\min}) \tag{3.2.4}$$

- Σ-*Minimum:*

$$\mathrm{Min}_\Sigma[f_{\boldsymbol{\nu}}] \geq (1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\,\lambda_{\max}) \tag{3.2.5}$$

- Σ-*Variance:*

$$\mathrm{Var}_\Sigma[f_{\boldsymbol{\nu}}] \leq \frac{(1-\gamma)^2 B_{\boldsymbol{\nu}}^2(\lambda_{\max}-\lambda_{\min})^2}{\gamma} \tag{3.2.6}$$

Proof. Consider the entropy bounding property (3.2.1) but with $\mathcal{C}$ restricted to the scores of $\mathcal{G}$, i.e., we have an affine function $g_{\boldsymbol{\nu}} : \mathfrak{P}_{\mathcal{G}} \to \mathbb{R}$ such that

$$g_{\boldsymbol{\nu}}(\boldsymbol{q}) \leq \inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i^{\mathrm{test}}(\sigma)_{\boldsymbol{C}_i(\mathcal{G})} = \tau_{\boldsymbol{q}}} H(A_iB_i|X_iY_iR')_{\mathcal{N}_i(\sigma)},$$

for all $\boldsymbol{q} \in \mathfrak{Q}_{\mathcal{G}}$.

As conditioning on additional side information will not increase the von Neumann entropy,[5] we may condition on whether or not the round was a test round,

$$\begin{aligned} H(A_iB_i|X_iY_iR')_{\mathcal{N}_i(\sigma)} &\geq H(A_iB_i|X_iY_iT_iR')_{\mathcal{N}_i(\sigma)} \\ &= \gamma H(A_iB_i|X_iY_i, T_i=1, R')_{\mathcal{N}_i(\sigma)} + (1-\gamma)H(A_iB_i|X_iY_i, T_i=0, R')_{\mathcal{N}_i(\sigma)} \\ &> (1-\gamma)H(A_iB_i|X_i=\tilde{x}, Y_i=\tilde{y}, T_i=0, R')_{\mathcal{N}_i(\sigma)} \end{aligned}$$

where in the final line we have used the fact that the inputs are fixed for generation rounds. As the min-entropy lower bounds the von Neumann entropy, we arrive at the bound

$$H(A_iB_i|X_iY_iR')_{\mathcal{N}_i(\sigma)} > (1-\gamma)H_{\min}(A_iB_i|X_i=\tilde{x}, Y_i=\tilde{y}, T_i=0, R')_{\mathcal{N}_i(\sigma)}.$$

Using the relaxed guessing probability program and its dual, we can lower bound the right-hand side. Specifically, for a single generation round

$$\begin{aligned} H_{\min}(AB|X=\tilde{x}, Y=\tilde{y}, T=0, R') &= -\log(p_{\mathrm{guess}}(\boldsymbol{q})) \\ &\geq -\log(\boldsymbol{\lambda}_{\boldsymbol{\nu}}^{(k)} \cdot \boldsymbol{q}), \end{aligned}$$

holds for all $k \in \mathbb{N}$, any $\boldsymbol{\nu} \in \mathfrak{Q}_{\mathcal{G}}^{(k)}$ and any quantum system realising the expected statistics $\boldsymbol{q} \in \mathfrak{Q}_{\mathcal{G}}$. In the final line we used the monotonicity of the logarithm together with the fact that a solution to the relaxed dual program, for any parameterization $\boldsymbol{\nu} \in \mathfrak{Q}_{\mathcal{G}}^{(k)}$, provides a linear function $\boldsymbol{q} \mapsto \boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{q}$ that is everywhere on $\mathfrak{Q}_{\mathcal{G}}^{(k)}$ greater than p_{guess}. Note that this bound is independent of the quantum state for which the entropy is evaluated and therefore automatically bounds the infimum. Dropping the (k) for notational ease, we may recover the desired affine property by taking a first order expansion about the point $\boldsymbol{\nu}$,

$$-(1-\gamma)\log(\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu}) - (1-\gamma)\sum_{c \in \mathcal{G}} \frac{\lambda_{\boldsymbol{\nu}}(c)}{\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu} \ln 2}(q(c) - \nu(c)).$$

[5] This is a consequence of the *strong subadditivity property* of the von Neumann entropy [46]. For any $\rho \in \mathcal{S}(ABC)$, we have

$$H(ABC) + H(B) \leq H(AB) + H(BC).$$

After rearranging this expression we find that $H(A|BC) \leq H(A|B)$.

Introducing, the constants $A_{\boldsymbol{\nu}} = \frac{1}{\ln 2} - \log(\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu})$ and $B_{\boldsymbol{\nu}} = \frac{1}{\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu} \ln 2}$ we may rewrite this expansion as the function

$$g_{\boldsymbol{\nu}}(\boldsymbol{q}) := (1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{q}),$$

which satisfies

$$g_{\boldsymbol{\nu}}(\boldsymbol{q}) \leq \inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i^{\text{test}}(\sigma)_{\boldsymbol{C}_i}(\mathcal{G}) = \tau_{\boldsymbol{q}}} H(A_i B_i | X_i Y_i R')_{\mathcal{N}_i(\sigma)},$$

for all $\boldsymbol{q} \in \mathfrak{Q}_{\mathcal{G}}$. The statement then follows from applying Lemma 3.1 to $g_{\boldsymbol{\nu}}$, noting $\text{Max}[g_{\boldsymbol{\nu}}] = (1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \lambda_{\min})$ and $\text{Min}[g_{\boldsymbol{\nu}}] = (1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \lambda_{\max})$. ■

Example 3.1. Taking the nonlocal game $\mathcal{G}_{\text{CHSH}}$ introduced in Example 2.4, we can use the above lemma to construct a min-tradeoff function. Fixing the probability of testing, $\gamma = 5 \times 10^{-3}$, we consider a device which behaves (during a test round) according to the expected frequency distribution $\boldsymbol{\omega} = (\omega_{\text{align}}, \omega_{\text{CHSH}}, 1 - \omega_{\text{align}} - \omega_{\text{CHSH}})$. In Fig. 3.1, we plot the certifiable min-entropy of a single generation round for a range of $\boldsymbol{\omega}$. We see that as the scores approach $\boldsymbol{\omega} = \frac{1}{2}\left(1, \frac{2+\sqrt{2}}{4}, \frac{2-\sqrt{2}}{4}\right)$, we are able to certify almost[6] two bits of randomness using $\mathcal{G}_{\text{CHSH}}$.

3.2.2 Application to the spot-checking protocol

After fixing the parameters of the protocol and constructing a min-tradeoff function $f_{\min}$, the user proceeds with the remaining steps of Protocol QRE: accumulation, evaluation and extraction. Recall that if the protocol does not abort, then with high probability the generated string $A_1^n B_1^n$ should contain at least some computable quantity of smooth min-entropy. The following lemma applies the EAT to deduce a lower bound on the amount of entropy produced by the devices.

Lemma 3.3 (Accumulated entropy)**.** *Let the randomness expansion procedure and all of its parameters be as defined in Fig. 3.3. Furthermore, let Ω be the event that the protocol does not abort (cf.* (3.1.2)*) and let $\rho_{|\Omega}$ be the final state of the system conditioned on Ω. Then, for any $\beta, \epsilon_s, \epsilon_{\text{EAT}} \in (0,1)$ and any choice of min-tradeoff function $f_{\boldsymbol{\nu}} \in \mathcal{F}_{\min}$, either Protocol QRE aborts with probability greater than $1 - \epsilon_{\text{EAT}}$ or*

$$H_{\min}^{\epsilon_s}(A_1^n B_1^n | X_1^n Y_1^n E)_{\rho_{|\Omega}} > (1-\gamma) n \left(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot (\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}_{\pm})\right) - n(\epsilon_V + \epsilon_K) - \epsilon_\Omega, \tag{3.2.7}$$

where

$$\epsilon_V := \frac{\beta \ln 2}{2}\left(\log\left(2|\mathcal{AB}|^2 + 1\right) + \sqrt{\frac{(1-\gamma)^2 B_{\boldsymbol{\nu}}^2 (\lambda_{\max} - \lambda_{\min})^2}{\gamma} + 2}\right)^2, \tag{3.2.8}$$

$$\epsilon_K := \frac{\beta^2}{6(1-\beta)^3 \ln 2} 2^{\beta(\log|\mathcal{AB}| + (1-\gamma)B_{\boldsymbol{\nu}}(\lambda_{\max} - \lambda_{\min}))} \ln^3\left(2^{\log|\mathcal{AB}| + (1-\gamma)B_{\boldsymbol{\nu}}(\lambda_{\max} - \lambda_{\min})} + e^2\right), \tag{3.2.9}$$

[6] Due to the infrequent testing we are actually only able to certify a maximum of $2 \cdot (1-\gamma)$ bits per interaction.

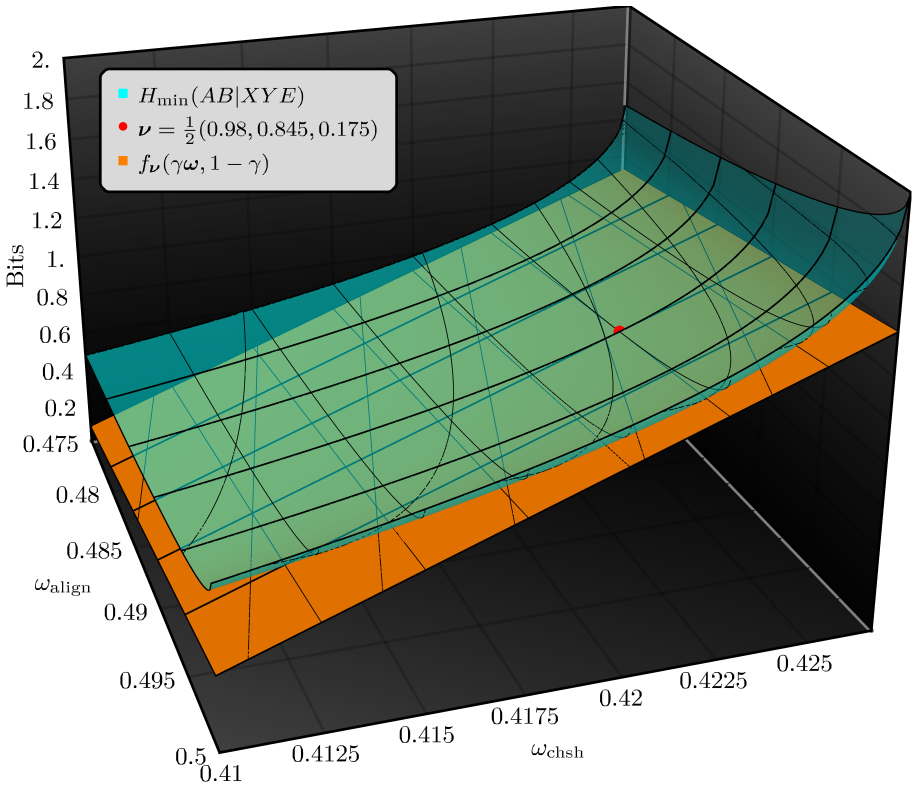


Figure 3.1: A plot of a lower bound on the certifiable min-entropy produced during a single round of the protocol. This lower bound was calculated using Prog. 2.7.13 relaxed to the second level of the NPA hierarchy. In addition, we plot a min-tradeoff function $f_{\boldsymbol{\nu}}$ evaluated for distributions of the form $\boldsymbol{p} = (\gamma\boldsymbol{\omega}, 1-\gamma)$ for $\boldsymbol{\omega} \in \Omega_{\mathcal{G}}$, i.e. expected frequency distributions over $\mathcal{G} \cup \{\perp\}$ that are compatible with the spot-checking structure of the rounds. Since $f_{\boldsymbol{\nu}}$ is the tangent plane to the surface at the point $\boldsymbol{\nu}$ it forms an affine lower bound on the min-entropy of any quantum distribution compatible with the protocol, i.e. any $\boldsymbol{q} \in \Omega_{\mathcal{G}}$ such that $\Sigma_{\boldsymbol{q}} \neq \emptyset$.

$$\epsilon_\Omega := \frac{1}{\beta}(1 - 2\log(\epsilon_{\text{EAT}}\epsilon_s)) \tag{3.2.10}$$

and $\boldsymbol{\delta}_\pm$ *is the vector with components* $\delta_i \cdot \text{sgn}(-\lambda_i)$.

Proof. Let $\{\mathcal{N}_i\}_{i\in[n]}$ be the set of channels implementing the entropy accumulation subprocedure of Protocol QRE. Comparing this procedure with the definition of the EAT channels Def. 2.5, we have $\mathcal{N}_i : \mathcal{S}(R_{i-1}) \to \mathcal{S}(A_iB_iX_iY_iT_iC_iR_i)$ with $A_i, B_i, X_i, Y_i, T_i, C_i$ finite dimensional classical systems, R_i an arbitrary quantum system and the score C_i is a deterministic function of the values of the other classical systems. Furthermore, the inputs to the protocol for the i^{th} round, (X_i, Y_i, T_i), are chosen independently of all other systems in the protocol and so the conditional independence constraints $I(A_1^{i-1}B_1^{i-1}{:}X_iY_i|X_1^{i-1}Y_1^{i-1}E) = 0$ hold trivially. The conditions necessary for $\{\mathcal{N}_i\}_{i\in[n]}$ to be EAT-channels are satisfied and

by Lemma 3.2 $f_{\boldsymbol{\nu}}$ is a min-tradeoff function for these channels. We now satisfy all of the prerequisites required to use the EAT.

Consider the pass probability of the protocol, $\mathbb{P}[\Omega]$. There are two possibilities: either, $\mathbb{P}[\Omega] < \epsilon_{\text{EAT}}$ in which case the protocol will abort with probability greater than $1-\epsilon_{\text{EAT}}$, or $\epsilon_{\text{EAT}} \leq \mathbb{P}[\Omega]$. In the latter case we can replace the unknown $\mathbb{P}[\Omega]$ in (2.3.15) with ϵ_{EAT} as this can only increase the error term ϵ_Ω. The EAT then asserts that

$$H_{\min}^{\epsilon_s}(A_1^n B_1^n | X_1^n Y_1^n E)_{\rho_{|\Omega}} > n \inf_{C_1^n \in \Omega} f_{\boldsymbol{\nu}}(F_{C_1^n}) - n(\epsilon_V + \epsilon_K) - \epsilon_\Omega,$$

for any choice of min-tradeoff function $f_{\boldsymbol{\nu}} \in \mathcal{F}_{\min}$.

As the min-tradeoff functions are affine, a lower bound on the infimum over region of possible scores specified by the success event,

$$\Omega = \{C_1^n \mid \gamma(\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}) < \boldsymbol{F}_{C_1^n}(\mathcal{G}) < \gamma(\boldsymbol{\omega}_{\mathcal{G}} + \boldsymbol{\delta})\},$$

can be readily computed. In particular, taking $\boldsymbol{p} = (\gamma(\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}_{\pm}), 1-\gamma)$ we have $f_{\boldsymbol{\nu}}(\boldsymbol{p}) \leq \inf_{C_1^n \in \Omega} f_{\boldsymbol{\nu}}(F_{C_1^n})$. Note that $\boldsymbol{p}$ may not correspond to a frequency distribution that could have resulted from a successful run of the protocol – it may not even be a probability distribution. However, it is sufficient for our purposes as an explicit lower bound on the infimum. Furthermore, noting that $f_{\boldsymbol{\nu}}(\boldsymbol{p}) = g_{\boldsymbol{\nu}}(\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}_{\pm})$, this lower bound may be written as

$$f_{\boldsymbol{\nu}}(\boldsymbol{p}) = (1-\gamma)\left(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot (\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}_{\pm})\right).$$

Inserting the min-tradeoff function properties: (3.2.4), (3.2.5) and (3.2.6); into the the EAT's error terms we get the explicit form of the quantities ϵ_V, ϵ_K and ϵ_Ω as seen above. ■

If the protocol does not abort during the accumulation procedure, the user may proceed by applying a quantum-proof strong extractor to the concatenated output string $A_1^n B_1^n$ resulting in a close to uniform bit-string of length approximately $(1-\gamma)n\left(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot (\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}_{\pm})\right) - n(\epsilon_V + \epsilon_K) - \epsilon_\Omega$.

Example 3.2. Continuing from Ex. 3.1, we look at the bound on the accumulated entropy specified by (3.2.7) for a range of choices of $f_{\boldsymbol{\nu}} \in \mathcal{F}_{\min}$. We consider a quantum implementation with an expected frequency distribution $\boldsymbol{\omega}_{\mathcal{G}} = (0.49, 0.4225, 0.0875)$, see the indicated point in Fig. 3.1. In Fig. 3.2 we see that our choice of min-tradeoff function can have a large impact on the quantity of entropy we are able to certify. The rough appearance of the EAT-rate surface is an artefact of obtaining local optima when we optimize over β. However, the plot gives some reassuring numerical evidence that, for the case of $\mathcal{G}_{\text{CHSH}}$, the certifiable randomness is continuous and concave in the family parameter $\boldsymbol{\nu}$.

The min-tradeoff function, $f_{\boldsymbol{\omega}_{\mathcal{G}}}$, certifies just under 0.939 bits per round. By applying a gradient-ascent algorithm we were able to improve this to 0.946 bits per round. In an attempt to avoid getting stuck within local optima we applied the algorithm several

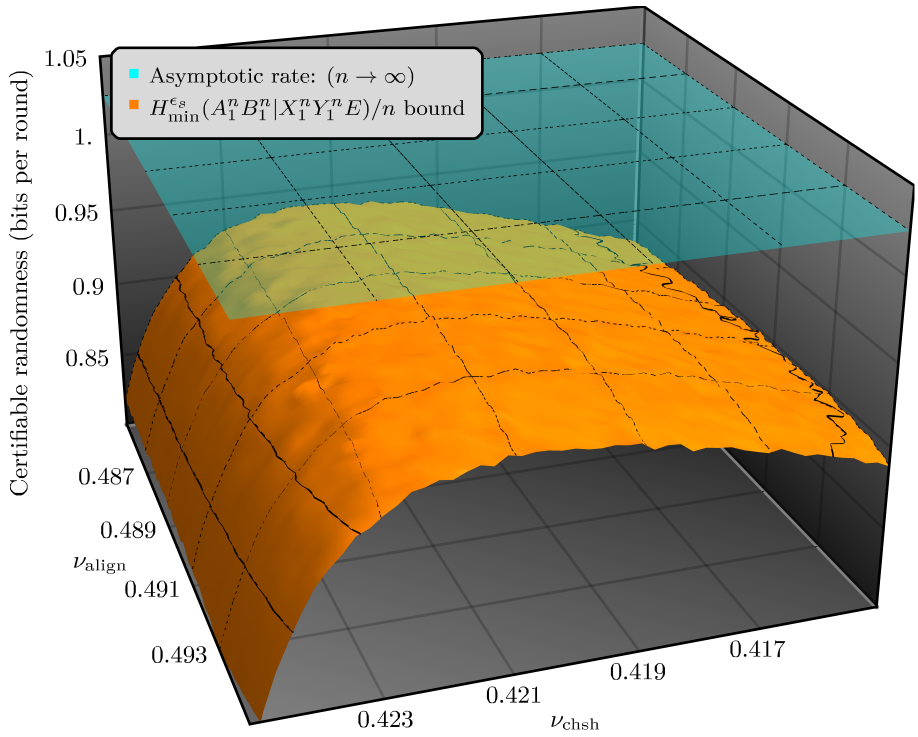


Figure 3.2: A plot of the randomness certification rate as we vary the min-tradeoff function. At each point $\boldsymbol{\nu}$ we evaluate the lower bound on $H_{\min}^{\epsilon_s}(A_1^n B_1^n | X_1^n Y_1^n E)/n$ as specified by (3.2.7) for the corresponding choice of min-tradeoff function $f_{\boldsymbol{\nu}}$, numerically optimizing the parameter β each time. The rough appearance of the surface results from finding local optima in the β optimization. For reference, we include a plot of the asymptotic min-entropy rate, i.e., the bound as $n \to \infty$, $\gamma \to 0$ and $\boldsymbol{\delta} \to \mathbf{0}$. The protocol parameters used during the calculations are: $n = 10^{10}$, $\gamma = 5 \times 10^{-3}$, $\delta_{\text{CHSH}} = \delta_{\text{align}} = 10^{-3}$ and $\epsilon_s = \epsilon_{\text{EAT}} = 10^{-8}$.

times, starting subsequent iterations at randomly chosen points close to the current optimum. The optimization led to an improved choice of min-tradeoff function $f_{\boldsymbol{\nu}^*}$, where $\boldsymbol{\nu}^* = (0.491, 0.421, 0.088)$.

3.2.3 Security of Protocol QRE

We refer to the pair of untrusted devices $\mathfrak{D}_{AB}$ as *honest* if during each interaction, the underlying quantum state shared amongst the devices and the measurements performed in response to inputs remain the same (i.e., the devices behave as the user expects). The following lemma provides a bound on the probability that an honest implementation of Protocol QRE aborts.

Lemma 3.4 (Completeness of Protocol QRE)**.** *Let Protocol QRE and all of its parameters be as defined in Fig. 3.3. Then, the probability that an honest implementation of Protocol QRE*

Protocol QRE

Parameters and notation:

$\mathfrak{D}_{AB}$ – a collection of two untrusted devices with inputs $\mathcal{X}$, $\mathcal{Y}$ and outputs $\mathcal{A}$, $\mathcal{B}$

$\mathcal{G} = (\mu, V)$ – a nonlocal game compatible with $\mathfrak{D}_{AB}$

$\boldsymbol{\omega}_{\mathcal{G}} \in \mathfrak{Q}_{\mathcal{G}}$ – an expected frequency distribution for $\mathcal{G}$

$\boldsymbol{\delta}$ – vector of confidence interval widths (satisfying $0 \leq \delta_k \leq \omega_k$ for all $k \in [|\mathcal{G}|]$)

$n \in \mathbb{N}$ – number of rounds

$\gamma \in (0,1)$ – probability of a test round

$(\tilde{x}, \tilde{y})$ – distinguished inputs for generation rounds

$f_{\min}$ – min-tradeoff function

$\epsilon_{\text{ext}} > 0$ – extractor error

$\epsilon_s \in (0,1)$ – smoothing parameter

$\epsilon_{\text{EAT}} \in (0,1)$ – entropy accumulation error

R_{ext} – quantum-proof $(k, \epsilon_{\text{ext}} + 2\epsilon_s)$-strong extractor

ℓ_{ext} – entropy loss induced by R_{ext}

Procedure:

1: Set $i = 1$.

2: **While** $i \leq n$:

Choose $T_i = 0$ with probability $1 - \gamma$ and otherwise $T_1 = 1$.

If $T_i = 0$:

Gen: Input $(\tilde{x}, \tilde{y})$ into the respective devices, recording the inputs $X_i Y_i$ and outputs $A_i B_i$. Set $C_i = \perp$ and $i = i + 1$.

Else:

Test: Play a single round of $\mathcal{G}$ on $\mathfrak{D}_{AB}$ using inputs sampled from μ, recording the inputs $X_i Y_i$ and outputs $A_i B_i$. Set $C_i = V(A_i B_i X_i Y_i)$ and $i = i + 1$.

3: Compute the empirical frequency distribution $\boldsymbol{F}_{C_1^n}$.

If $\gamma(\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}) < \boldsymbol{F}_{C_1^n}(\mathcal{G}) < \gamma(\boldsymbol{\omega}_{\mathcal{G}} + \boldsymbol{\delta})$:

Ext: Apply a strong quantum-proof randomness extractor R_{ext} to the output string $A_1^n B_1^n$ producing $f_{\min}(\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}_{\pm}) - \ell_{\text{ext}}$ bits $(\epsilon_{\text{ext}} + 2\epsilon_s)$-close to uniformly distributed.

Else:

Abort: Abort the protocol.

Figure 3.3: Template quantum-secure randomness expansion protocol

aborts is no greater than ε_{comp} *where*

$$\varepsilon_{comp} = 2 \sum_{k=1}^{|\mathcal{G}|} e^{-\frac{\gamma \delta_k^2}{3\omega_k} n}. \tag{3.2.11}$$

Proof. During the parameter estimation step of Protocol QRE, the protocol aborts if the observed frequency distribution $F_{C_1^n}$ fails to satisfy

$$\gamma(\boldsymbol{\omega}_{\mathcal{G}} - \boldsymbol{\delta}) < F_{C_1^n}(\mathcal{G}) < \gamma(\boldsymbol{\omega}_{\mathcal{G}} + \boldsymbol{\delta}).$$

Writing $F_{C_1^n}(\mathcal{G}) = (r_k)_{k=1}^{|\mathcal{G}|}$, $\boldsymbol{\omega}_{\mathcal{G}} = (\omega_k)_{k=1}^{|\mathcal{G}|}$ and $\boldsymbol{\delta} = (\delta_k)_{k=1}^{|\mathcal{G}|}$, the probability that an honest implementation of the protocol aborts can be written as

$$\mathbb{P}\left[\Omega^c\right] = \mathbb{P}\left[\bigcup_{k=1}^{|\mathcal{G}|}\left\{\left|r_k - \gamma\omega_k\right| \geq \gamma\delta_k\right\}\right] \leq \sum_{k=1}^{|\mathcal{G}|}\mathbb{P}\left[\left|r_k - \gamma\omega_k\right| \geq \gamma\delta_k\right].$$

Restricting ourselves to a single element r_k of $F_{C_1^n}(\mathcal{G})$, we can model its final value as the binomially distributed random variable $r_k \sim \frac{1}{n}\mathrm{Bin}\left(n, \gamma\omega_k\right)$. As a consequence of the Chernoff bound (see Corollary 2.1) and $\delta_k \leq \omega_k$, we have

$$\mathbb{P}\left[\left|r_k - \gamma\omega_k\right| \geq \gamma\delta_k\right] \leq 2\mathrm{e}^{-\frac{\gamma\delta_k^2 n}{3\omega_k}}.$$

Applying this bound to each element of the sum individually, we arrive at (3.2.11). ■

Remark 3.5. The completeness error in the above lemma only considers the possibility of the protocol aborting during the parameter estimation stage. However, if the initial random seed is a limited resource then this may pose an additional restriction on the protocol. In Lemma 3.7 we analyse the probability of failure of a specific algorithm for sampling the inputs of Protocol QRE. If required, the probability of failure for that algorithm could be incorporated into the completeness error.

Lemma 3.5 (Soundness of Protocol QRE)**.** *The soundness error of Protocol QRE is*

$$\varepsilon_{sound} = \max(\epsilon_{ext} + 2\epsilon_s, \epsilon_{\mathrm{EAT}}).$$

Proof. Recall from (3.1.3) that the soundness error is an upper bound on $\frac{1}{2}\mathbb{P}[\Omega]\cdot\|\rho_{ZE} - \tau_m \otimes \rho_E\|_1$. In the case $\mathbb{P}[\Omega] \leq \epsilon_{\mathrm{EAT}}$, we have $\frac{1}{2}\mathbb{P}[\Omega]\cdot\|\rho_{ZE} - \tau_m \otimes \rho_E\|_1 \leq \epsilon_{\mathrm{EAT}}$.

In the case $\mathbb{P}[\Omega] > \epsilon_{\mathrm{EAT}}$, Lemma 3.3 gives a bound on the accumulated entropy. Combining this with the definition of a quantum-proof strong extractor (Def. 2.7) and noting that the norm is non-increasing under partial trace (tracing out the extractor's seed) we obtain $\frac{1}{2}\mathbb{P}[\Omega]\cdot\|\rho_{ZE} - \tau_m \otimes \rho_E\|_1 \leq \epsilon_{\mathrm{ext}} + 2\epsilon_s$, from which the claim follows. ■

Remark 3.6. By choosing parameters such that $\epsilon_{\mathrm{EAT}} \leq \epsilon_{\mathrm{ext}} + 2\epsilon_s$ we can take the soundness error to be $\epsilon_{\mathrm{ext}} + 2\epsilon_s$.

Combining all of the previous results we arrive at the full security statement concerning Protocol QRE.

Theorem 3.1 (Security of Protocol QRE)**.** *Protocol QRE is an* $(\varepsilon_{comp}, \varepsilon_{sound})$*-secure randomness expansion protocol producing*

$$((1-\gamma)(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\boldsymbol{\lambda}_{\boldsymbol{\nu}}\cdot(\boldsymbol{\omega} - \boldsymbol{\delta}_{\pm})) - \epsilon_V - \epsilon_K)n - \epsilon_\Omega - \ell_{ext} \tag{3.2.12}$$

random bits at least ε_{sound}*-close to uniformly distributed, where* ε_{comp}*,* ε_{sound} *are given as in Lemma 3.4 (cf. Remark 3.5) and Lemma 3.5.*

Remark 3.7. The expected number of uniformly random bits required to execute Protocol QRE is $d \approx \big(\gamma H(\mu) + h(\gamma)\big) n$ (cf. Lemma 3.7).

Example 3.3. In Ex. 3.1 (cf. Fig. 3.1) and Ex. 3.2 (cf. Fig. 3.2) we used the protocol parameters: $n = 10^{10}$, $\gamma = 5 \times 10^{-3}$, $\delta(c) = 10^{-3}$ for each $c \in \mathcal{G}$ and $\epsilon_s = \epsilon_{\text{EAT}} = 10^{-8}$. The resulting implementation of Protocol QRE, using the nonlocal game $\mathcal{G}_{\text{CHSH}}$ with an expected distribution over the scores $\boldsymbol{\omega}_{\mathcal{G}} = (0.49, 0.4225, 0.0875)$, exhibits the following statistics.

Quantity	**Value**
Total accumulated entropy before extraction (no abort)	9.46×10^9
Expected length of required seed before extraction	5.54×10^8
Expected net-gain in entropy (no abort)	$8.91 \times 10^9 - \ell_{\text{ext}}$
Completeness error ($\varepsilon_{\text{comp}}$)	8.77×10^{-8}

3.3 Bookkeeping

Let us now investigate the amount of initial randomness required to run protocols within the framework. This supply of random bits is necessary for selecting the devices' inputs and seeding the extractor. In the forthcoming analysis we focus our attention on the former since the latter is dependent on the choice of extractor. Moreover, if one chooses to use a strong extractor, then the seed acts in a catalytic manner and thus, in this sense, can be regarded as free. Therefore, we restrict our considerations to the process of converting a uniform private seed into the device inputs required for running Protocol QRE. We begin by introducing an efficient algorithm for simulating the sampling of a target random variable T by sampling another random variable S [94]. We then apply this algorithm to the spot-checking protocol and use it to bound the size of the seed required to run Protocol QRE.

3.3.1 The interval algorithm

Let S and T be random variables taking values from their respective alphabets $\mathcal{S}$ and $\mathcal{T}$. The *interval algorithm* proposes a method by which we can use repeated samples of S to simulate a sample of T.

The distributions of the random variables S and T both form a partition of the unit interval: that is, to each outcome $s \in \mathcal{S}$ ($t \in \mathcal{T}$) we associate a subinterval of length $\mathbb{P}[S = s]$ ($\mathbb{P}[T = t]$). Similarly, if we repeatedly sample S, i.e. sample the product distribution S^k for some $k \in \mathbb{N}$, then this defines another, more fine-grained, partition of the unit interval. To execute the interval algorithm we repeatedly sample S, recording the outcomes $\boldsymbol{s} = (s_1, s_2, \dots)$ until the interval corresponding to our sequence of outcomes is contained entirely within one of the intervals defined by the distribution of the random variable T. Once this termination criterion is met, the algorithm returns the label $t \in \mathcal{T}$ of the interval that our sequence is entirely contained within.

Intuitively, the algorithm can be seen to converge, since, after k samples of S, S^k defines a fine-grained partition of the unit interval with $|\mathcal{S}|^k$ subintervals, the largest of which has size $(\max_{s\in\mathcal{S}}\mathbb{P}[S=s])^k$. Considering then the subinterval corresponding to an outcome t of T, as k becomes large the subintervals of S^k contained entirely inside the subinterval corresponding to t have a combined length that is close to $\mathbb{P}[T=t]$ with an error that decreases exponentially in k. As this holds for all $t\in\mathcal{T}$, for large enough k the algorithm returns t with probability that is close to $\mathbb{P}[T=t]$ and thus the procedure provides a good approximation for sampling T.

For simplicity we shall now restrict ourselves to the case where S corresponds to sampling a uniformly random bit. The question remains as to how efficient this procedure is in terms of the number of random bits required. Denoting the length of seed required for the interval algorithm to terminate by N, we have, by Theorem 3 in [94],

$$\mathbb{E}[N]\leq H(T)+3. \tag{3.3.1}$$

The sampling procedure defined by the interval algorithm does not bound the maximum value that N can take (although the probability that the algorithm does not terminate decreases exponentially in the number of samples). In order to produce large deviation bounds on the number of bits required to execute our protocol, we place an upper limit on the maximum seed length. We thus propose an adapted sampling procedure which we call the *rounded interval algorithm* (RIA), which forcefully terminates if the seed length reaches the upper bound of $k_{\max}$ bits. Should the RIA fail to terminate after $k_{\max}$ steps, then the output sequence generated will correspond to the subinterval $I(r)=\left[\frac{r}{2^{k_{\max}}},\frac{r+1}{2^{k_{\max}}}\right)$, for some $r\in\{0,1,\ldots,2^{k_{\max}}-1\}$, that is not entirely contained within one of the subintervals defined by T. If this occurs, we *round down*: selecting the interval I_t for which $\frac{r}{2^{k_{\max}}}\in I_t$.

Remark 3.8. The rounding procedure bounds the maximum seed length as $N\leq k_{\max}$ and therefore, the inequality (3.3.1) also holds for the RIA.

The truncation of the interval algorithm, described in the RIA, does not significantly hinder the convergence of the procedure as the following lemma shows.

Lemma 3.6. *Let T be a random variable taking values in some alphabet $\mathcal{T}$. Let T' be the distribution sampled using the RIA with target distribution T. Then*

$$\Delta(T,T')\leq|\mathcal{T}|\,2^{-(k_{\max}+1)},$$

where $k_{\max}$ is the maximum number of input bits that can be used by the RIA.

Proof. Consider the partitions of the unit interval $\{I(t)\}_{t\in\mathcal{T}}$ and $\{I'(t)\}_{t\in\mathcal{T}}$ corresponding to the distributions p_T and $q_{T'}$ of T and T' respectively. The intervals of T' take the form

$$I'(t)=\bigcup_r\left[\frac{r}{2^{k_{\max}}},\frac{r+1}{2^{k_{\max}}}\right)$$

where the (potentially empty) union is taken over all $r \in \mathbb{N}_0$ such that $r2^{-k_{\max}} \in I(t)$. The intervals within the union are either contained fully within the corresponding outcome interval of T, i.e., $\left[\frac{r}{2^{k_{\max}}}, \frac{r+1}{2^{k_{\max}}}\right) \subseteq I(t)$, or they are included as a result of rounding. Thus we may write

$$|I'(t)| = |\{r \mid r2^{-k_{\max}} \in I(t), r \in \mathbb{N}_0\}|2^{-k_{\max}}.$$

By a straightforward counting argument, there are at least $\lfloor |I(t)|2^{k_{\max}} \rfloor$ such values of r, and at most $\lceil |I(t)|2^{k_{\max}} \rceil$. We hence have

$$|I(t)|2^{k_{\max}} - 1 \leq |I'(t)|2^{k_{\max}} \leq |I(t)|2^{k_{\max}} + 1,$$

and therefore

$$|p_T(t) - p_{T'}(t)| \leq 2^{-k_{\max}},$$

holds for all $t \in \mathcal{T}$. Applying this bound to each term within the $\Delta(T, T')$ sum completes the proof. ■

3.3.2 Input randomness for Protocol QRE

Following Protocol QRE, we look to use the RIA to sample the devices' inputs for each round. In adherence with the conditional independence constraints of Def. 2.5, the natural procedure would be to sample at the beginning of each round. However, sampling many inputs at once turns out to be much more efficient in terms of the length of seed required. This can be seen by considering the bound (3.3.1) together with the property $H(T^n) = nH(T)$: by sampling the joint distribution, T^n, the expected saving is about $3n$ bits when compared to repeating a single sample n times. This is significant as the bound on the entropy we accumulate also grows linearly in n.

Fortunately, this joint sampling can be implemented while maintaining the conditional independence assumption required for the EAT analysis. Within the assumptions of Protocol QRE we allow the honest parties access to a trusted classical computer, which also contains some trusted data storage—we assume that the parties can record their output strings without leakage. Thus, the honest parties may select the devices' inputs prior to the device interaction phase, store them securely within the classical computer and, at the beginning of each round, feed the corresponding inputs to their devices. In such a scenario we retain the conditional independence assumption specified in Def. 2.5. Due to potential computational constraints we will not assume that all n rounds are sampled at once. Instead, we split the n rounds into at most $\lceil n/m \rceil$ blocks of size m and apply the RIA to sample the inputs of each block separately. For simplicity, we assume that $n/m \in \mathbb{N}$ and henceforth remove the ceiling function from the analysis.

Recall that for the i^{th} round, the user first uses T_i to decide whether the round is a test round, and, if so, they choose inputs according to the input distribution μ of their chosen nonlocality test. Otherwise, if $T_i = 0$, they supply their devices with the fixed inputs $\tilde{x}$ and

$\tilde{y}$. The probability mass function of the joint random variables $X_i Y_i T_i$, representing the i^{th} round's inputs, is therefore

$$\mathbb{P}[(X_i, Y_i, T_i) = (x_i, y_i, t_i)] = \begin{cases} \gamma\,\mu(x,y) & \text{for } (x_i, y_i, t_i) = (x, y, 1), \\ (1-\gamma) & \text{for } (x_i, y_i, t_i) = (\tilde{x}, \tilde{y}, 0) \\ 0 & \text{otherwise} \end{cases}. \tag{3.3.2}$$

Following (3.3.1), if M is the seed length required to sample one of the m blocks of rounds, then we have

$$\mathbb{E}[M] \leq \frac{\big(\gamma H(\mu) + h(\gamma)\big)\, n}{m} + 3 \tag{3.3.3}$$

where $H(\mu)$ is the Shannon entropy of the distribution μ and $h(\cdot)$ is the binary entropy.

The following lemma gives a probabilistic bound on the total length of the random seed required to sample the inputs for the devices.

Lemma 3.7. *Let the parameters of Protocol QRE be as defined in Fig. 3.3 and let $k_{\max} \in \mathbb{N}$ be the maximum permitted seed length for an instance of the RIA. Then, with probability greater than $(1 - \epsilon_{RIA})$, we can use m instances of the RIA to simulate the sampling of every device input required to execute Protocol QRE with a uniform seed of length no greater than $N_{\max}$, where*

$$N_{\max} = 2\kappa \tag{3.3.4}$$

$$\epsilon_{RIA} = e^{-2\kappa^2 / m k_{\max}^2} \tag{3.3.5}$$

and $\kappa = \big(\gamma H(\mu) + h(\gamma)\big)\, n + 3m$. Moreover, the sampled distribution lies within a statistical distance of

$$\epsilon_{dist} = m\, 2^{n \log(\text{supp}(\mu)+1)/m - (k_{\max}+1)}, \tag{3.3.6}$$

from the target distribution, where $\text{supp}(\mu) := |\{(x,y) \in \mathcal{X}\mathcal{Y} \mid \mu(x,y) > 0\}|$.

Proof. Consider the sequence $(M_i)_{i=1}^m$ of i.i.d. random variables representing the number of random bits required to choose the inputs for the i^{th} block and the corresponding random sum $N = \sum_{i=1}^m M_i$. By (3.3.3), the expected number of bits required to select all of the inputs for the protocol can be bounded above by $\kappa = \big(\gamma H(\mu) + h(\gamma)\big)\, n + 3m$. Using Hoeffding's inequality Lemma 2.2, we can bound the probability that N greatly exceeds this value,

$$\mathbb{P}[N \geq \kappa + t] \leq e^{-2t^2 / m k_{\max}^2},$$

for some $t > 0$. Setting $t = \kappa$ this becomes

$$\mathbb{P}[N \geq 2\kappa] \leq e^{-2\kappa^2 / m k_{\max}^2}.$$

Although κ is not exactly the expected value of N, which is the quantity appearing in Hoeffding's bound, the bound still holds as $\kappa \geq \mathbb{E}[N]$.

It remains to bound the statistical distance between the sampled random variable $\hat{I}_1^m = (\hat{X}_1^m, \hat{Y}_1^m, \hat{T}_1^m)$ and the target random variable $I_1^m = (X_1^m, Y_1^m, T_1^m)$. For each block of rounds, the corresponding random variable I_i can take one of a possible $(\text{supp}(\mu)+1)^{n/m}$ different values. Therefore, by Lemma 3.6, we have for the i^{th} block of rounds

$$\begin{aligned}\Delta(I_i, \hat{I}_i) &\leq (\text{supp}(\mu)+1)^{n/m} 2^{-(k_{\max}+1)} \\ &= 2^{n\log(\text{supp}(\mu)+1)/m-(k_{\max}+1)}\end{aligned}$$

Since $\Delta(P,Q)$ is a metric and hence satisfies the triangle inequality [51], the statistical distance between independently repeated samples can grow no faster than linearly, i.e., $\Delta(I^m, I'^m) \leq m\Delta(I, I')$.[7] This completes the proof. ■

[7] More specifically, we can iteratively apply the triangle inequality as follows

$$\begin{aligned}\Delta(P^n, Q^n) &\leq \Delta(P^n, P^{n-1}Q) + \Delta(P^{n-1}Q, Q^n) \\ &= \Delta(P,Q) + \Delta(P^{n-1}, Q^{n-1}).\end{aligned}$$

CHAPTER 4

NOISE ROBUSTNESS OF THE FRAMEWORK AND FEASIBILITY OF RANDOMNESS EXPANSION

This chapter looks at some explicit constructions of randomness expansion protocols. We model their implementation on systems of entangled qubits and compare their respective randomness expansion rates when exposed to inefficient detectors. Towards the end of the chapter, we also investigate the question of whether net-positive randomness expansion rates can be achieved using current technology.

4.1 Additional nonlocality tests

We begin by introducing two different nonlocality tests that will be included in the constructions alongside the extended CHSH game $\mathcal{G}_{\text{CHSH}}$ (cf. Example 2.4). The first game, which we refer to as the *empirical behaviour game*, provides the strongest possible device-independent characterization of the untrusted devices.

Empirical behaviour game ($\mathcal{G}_{\text{EB}}$). The *empirical behaviour game* ($\mathcal{G}_{\text{EB}}$) is a nonlocal game that estimates the underlying behaviour of $\mathfrak{D}_{AB}$, i.e., it attempts to characterise each individual probability $p(a,b|x,y)$. We may construct this by associating with each input-output tuple $(a,b,x,y) \in \mathcal{AB}\mathcal{X}\mathcal{Y}$ a corresponding score $c_{ab|xy} \in \mathcal{G}$ and defining the scoring rule

$$V_{\text{EB}}(a,b,x,y) := c_{ab|xy},$$

for each $(a,b,x,y) \in \mathcal{AB}\mathcal{X}\mathcal{Y}$. Then, for any input distribution μ_{EB} with full support on the alphabets $\mathcal{X}\mathcal{Y}$, the collection $\mathcal{G}_{\text{EB}} = (\mu_{\text{EB}}, V_{\text{EB}})$ forms a nonlocal game. Moreover, for agents playing according to some strategy $\boldsymbol{p} \in \mathfrak{Q}$, the expected frequency distribution over the

scores is precisely the joint distribution,

$$\begin{aligned}\omega_{\mathrm{EB}}(a,b,x,y) &= \mu_{\mathrm{EB}}(x,y)p(a,b|x,y)\\ &= p(a,b,x,y).\end{aligned}$$

As $\mathcal{G}_{\mathrm{EB}}$ can be defined for any collection of input-output alphabets, we indicate the size of these alphabets as superscripts, i.e., $\mathcal{G}_{\mathrm{EB}}^{|\mathcal{X}||\mathcal{Y}||\mathcal{A}||\mathcal{B}|}$.[1]

Remark 4.1. The scoring rule for $\mathcal{G}_{\mathrm{EB}}$, as defined above, has several redundant components, see Fig. 4.1. In fact, there are only $[(|\mathcal{A}|-1)|\mathcal{X}|+1][(|\mathcal{B}|-1)|\mathcal{Y}|+1]-1$ free parameters [95]. Knowing this we can reduce the number of scores in our nonlocal game and, in turn, the number of constraints we impose in our SDPs.[2] Using the table presented in Fig. 4.1, we can associate a score with every element that is not coloured. For the coloured elements we can assign a score c_{norm} which normalizes the resulting frequency distribution.

In practice we are limited to collections of finite statistics and so we face a tradeoff between how fine-grained of a characterization of our devices we choose to pursue and how confident we are that our observations are not skewed by statistical fluctuations. As a consequence, we may be required to collect substantially more test data if we want to use nonlocality tests with larger score alphabets. The joint correlators game, which we now introduce, offers an intermediate step between $\mathcal{G}_{\mathrm{CHSH}}$ and $\mathcal{G}_{\mathrm{EB}}$.

Joint correlators game ($\mathcal{G}_{\langle AB\rangle}$)**.** Specifically, for each $(x,y)\in\mathcal{XY}$ we define a score c_{xy} and a scoring rule

$$V_{\langle AB\rangle}(a,b,x,y) := \begin{cases} c_{xy} & \text{if } a=b\\ c_{\mathrm{norm}} & \text{otherwise.}\end{cases}$$

That is, for a pair of inputs (x,y) the score is recorded as c_{xy} whenever the agents' outcomes agree. Otherwise, they record some normalization score c_{norm}. The input distribution can then be specified as one sees fit: we shall use the uniform distribution over $\mathcal{XY}$. We refer to this game by the symbol $\mathcal{G}_{\langle AB\rangle}$ and as before we will indicate the sizes of the input-output alphabets as superscripts.

4.2 Comparison of protocols on noisy qubit systems

Let us now introduce a qubit implementation of the protocols that we will use to analyse the noise robustness of the framework. We retain the protocol parameter choices from the previous examples: $n=10^{10}$, $\gamma=5\times10^{-3}$ and $\epsilon_s=\epsilon_{\mathrm{EAT}}=10^{-8}$, except we now set the

[1] As we consider only binary output alphabets, we will not include their sizes in the superscript, i.e., we will write $\mathcal{G}_{\mathrm{EB}}^{23}$ instead of $\mathcal{G}_{\mathrm{EB}}^{2322}$.

[2] It is important to remove redundant constraints in practice as they can lead to numerical instabilities.

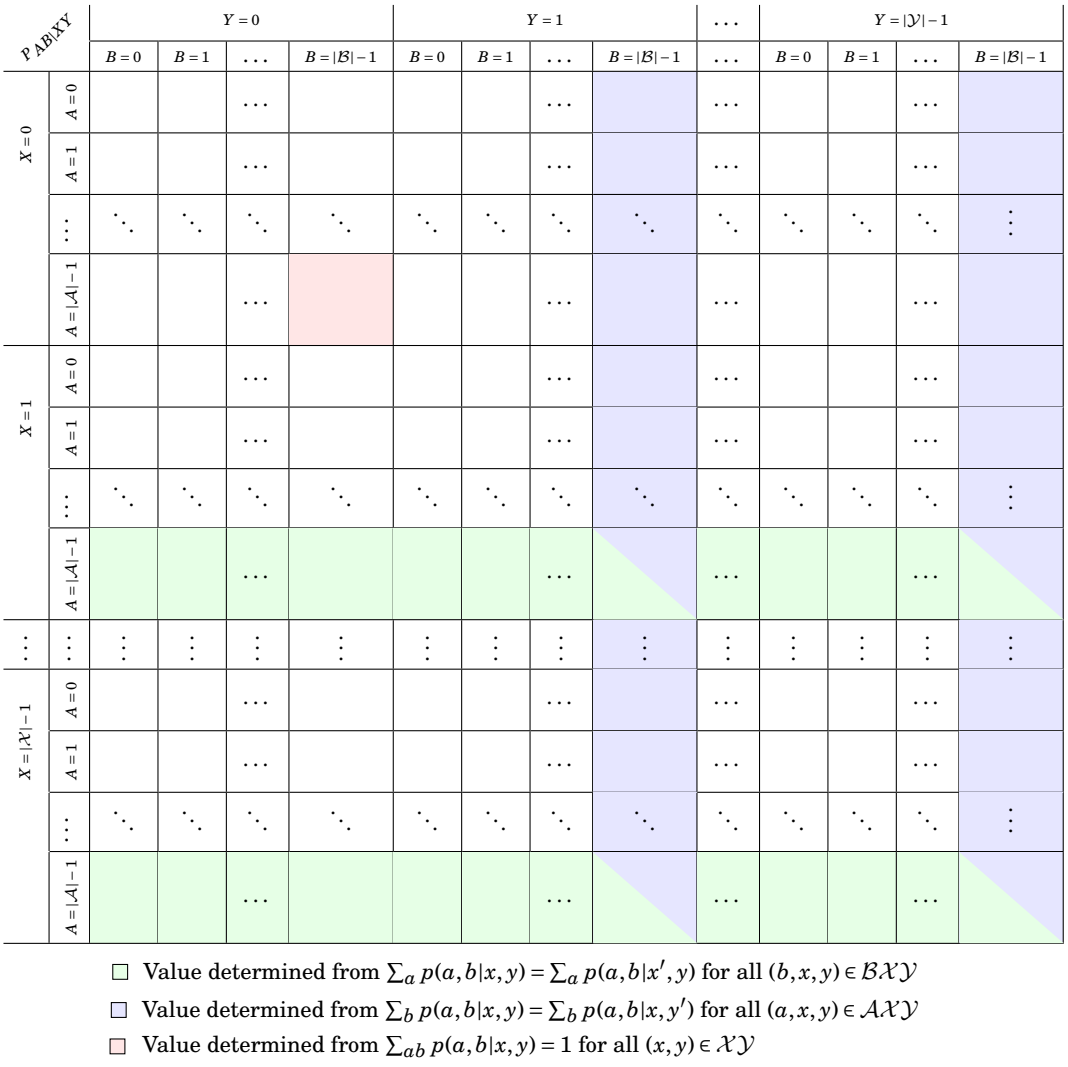


Figure 4.1: Table showing redundant elements of a no-signalling distribution $\boldsymbol{p} \in \mathfrak{N}$.

confidence interval width parameter to

$$\delta_k = \sqrt{\frac{3\,\omega_k \ln(2/\varepsilon_{\text{comp}})}{\gamma n}}, \tag{4.2.1}$$

in order to have a similar completeness error $\varepsilon_{\text{comp}} \approx 10^{-12}$ across the different protocols.[3] We suppose that the joint state of the devices at the start of each round is given by a pure, non-maximally entangled state of the form

$$|\psi(\theta)\rangle_{AB} = \cos(\theta)|00\rangle + \sin(\theta)|11\rangle\,, \tag{4.2.2}$$

for $\theta \in (0, \pi/2)$. We denote the corresponding density operator by $\rho_\theta = |\psi(\theta)\rangle\langle\psi(\theta)|$. For simplicity we restrict to projective measurements within the x-z plane of the Bloch-sphere,

[3] In practice one would fix the soundness error of the protocol. However, because the soundness error is also dependent on the extraction phase we instead assume independence of rounds and fix the completeness error.

i.e., measurements $\{\Pi(\varphi), \mathbb{1} - \Pi(\varphi)\}$, where

$$\Pi(\varphi) = \begin{pmatrix} \cos^2(\varphi/2) & \cos(\varphi/2)\sin(\varphi/2) \\ \cos(\varphi/2)\sin(\varphi/2) & \sin^2(\varphi/2) \end{pmatrix} \tag{4.2.3}$$

for $\varphi \in (0, 2\pi]$. We denote the projectors associated with the j^{th} outcome of the i^{th} measurement by $A_{j|i}$ and $B_{j|i}$. The elements of the devices' behaviour can then be written as

$$p(a,b|x,y) = \operatorname{Tr}\left[\rho_\theta (A_{a|x} \otimes B_{b|y})\right]. \tag{4.2.4}$$

Noise within a randomness expansion protocol may come from several different sources: within the creation and transmission of the states, as well as during the measurement process. Whilst one can use heralding to account for losses incurred during state transmission, losses that occur within the secure laboratories (i.e. during the measurement process) cannot be ignored without opening a detection loophole (cf. Sec. 2.5.2). Inefficient detectors are a major contributor to the total experimental noise, so robustness to inefficient detectors is a necessary property for any practical randomness expansion protocol. We characterize detection efficiency by a single parameter $\eta \in (0, 1]$, representing the (independent) probability with which a measurement device successfully measures a received state and outputs the result.[4] We deal with failed measurements by assigning the outcome 0. Combining this with (4.2.4), we may write the behaviour as

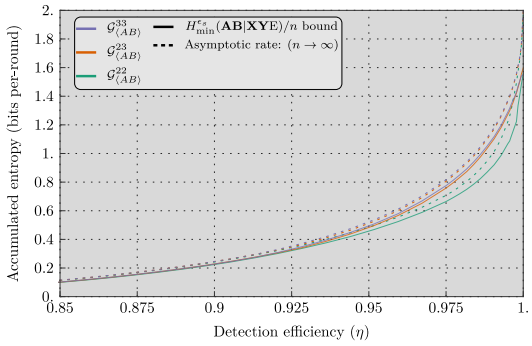
$$\begin{aligned} p(a,b|x,y) = \eta^2 \operatorname{Tr}\left[\rho_\theta (A_{a|x} \otimes B_{b|y})\right] + (1-\eta)^2 \delta_{0a}\delta_{0b} \\ + \eta(1-\eta)\left(\delta_{0a} \operatorname{Tr}\left[\rho_\theta (\mathbb{1} \otimes B_{b|y})\right] + \delta_{0b} \operatorname{Tr}\left[\rho_\theta (A_{a|x} \otimes \mathbb{1})\right]\right). \end{aligned} \tag{4.2.5}$$

For each protocol we consider lower bounds on two quantities: the min-entropy produced from a single interaction (before applying the EAT), $H_{\min}(AB|XYE)$, and the *EAT-rate*, $H^{\epsilon_s}_{\min}(A_1^n B_1^n | X_1^n Y_1^n E)/n$. The former quantity, which we refer to as the *asymptotic rate*, represents the maximum accumulation rate achievable with our numerical technique. It is a lower bound on $H^{\epsilon_s}_{\min}(A_1^n B_1^n | X_1^n Y_1^n E)/n$, specified by (3.2.7), as $n \to \infty$ and $\gamma, \delta \to 0$.[5] Comparing the asymptotic rate with the EAT-rate gives us a clear picture of the amount of entropy that we lose due to the effect of finite statistics.

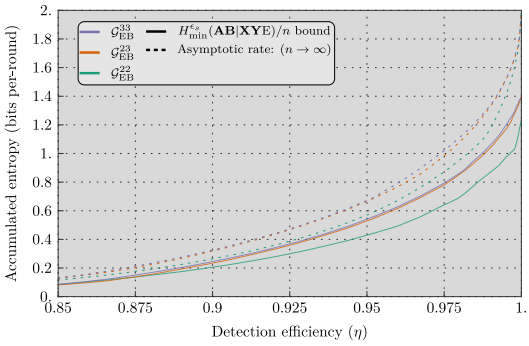
With inefficient detectors, partially entangled states can exhibit larger Bell-inequality violations than maximally entangled states [96]. To account for this we optimize both the state and measurement angles at each data point using the iterative optimization procedure detailed in [97]. We relax all programs to the second level of the NPA hierarchy and solve the resulting SDPs with the SDPA solver [84].

[4] For simplicity, we make the additional assumption that the detection efficiencies are constant amongst all measurement devices used within the protocol.

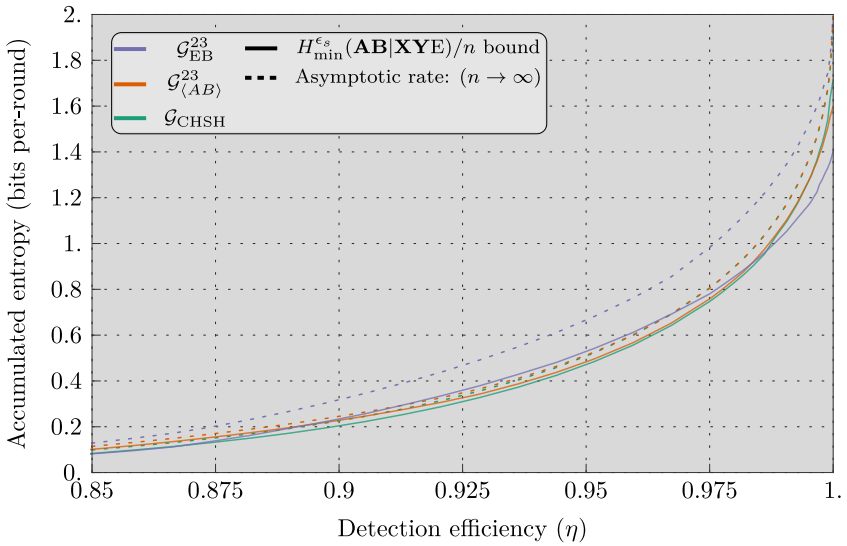
[5] In principle, we would rather characterise $H(AB|XYE)$ and the corresponding EAT-rate derived from it. However, in general we don't have suitable techniques to access these quantities in a device-independent manner.



(a) Comparison of $\mathcal{G}_{\langle AB\rangle}$ protocols



(b) Comparison of $\mathcal{G}_{\mathrm{EB}}$ protocols



(c) Comparison of protocols in the (2,3)-scenario

Figure 4.2: A plot of the asymptotic and EAT-rates for protocols using the nonlocal game families $\mathcal{G}_{\langle AB\rangle}$, $\mathcal{G}_{\mathrm{EB}}$ and $\mathcal{G}_{\mathrm{CHSH}}$.

In Fig. 4.2a and Fig. 4.2b we see that in both families of protocols considered, an increase in the number of inputs leads to higher rates. This increase is significant when moving from the (2,2)-scenario to the (2,3)-scenario. However, continuing this analysis for higher numbers of inputs we find that any further increases appear to have negligible impact on the overall robustness of the protocol.[6] Whilst all of the protocols achieve asymptotic rates of 2 bits per round when $\eta = 1$,[7] their respective EAT-rates at this point differ substantially. In Fig. 4.2c we see a direct comparison between protocols from the different families. The plot shows that, as expected, entropy loss is greater when using the nonlocality test G^{23}_{EB} as opposed to the other protocols. In particular, for high values of η we find that we are able

[6] This could also be an artefact of the assumed restriction to qubit systems.

[7] More precisely, the asymptotic rates are $2(1-\gamma)$ bits per round.

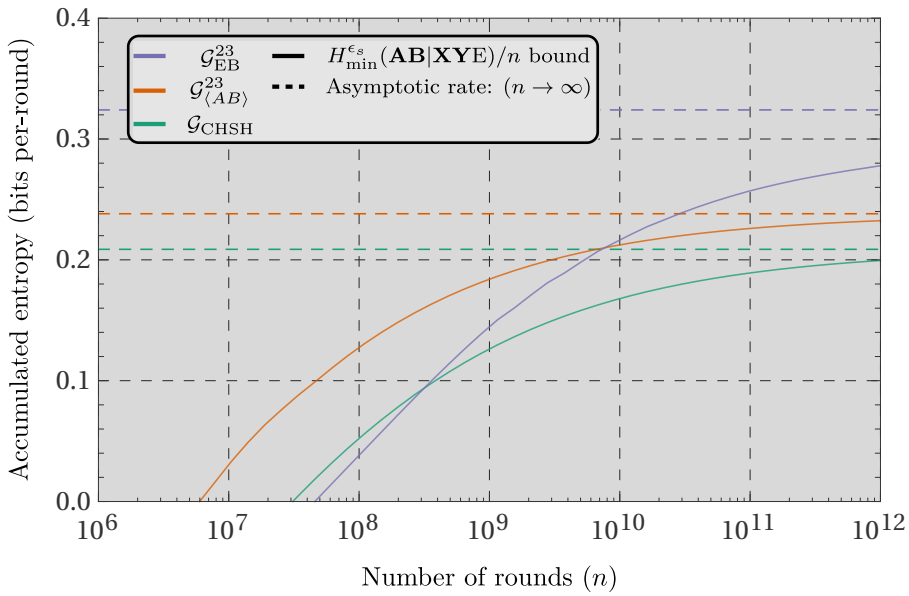


Figure 4.3: Comparison of the EAT-rates (cf. (3.2.7)) and their convergence to the asymptotic rates for protocols based on different nonlocality tests. The rates were derived by assuming a qubit implementation of the protocols with a detection efficiency of $\eta = 0.9$, optimizing the state and measurement angles in order to maximise the asymptotic rate. Then, for each value of n we optimized the min-tradeoff function and β parameter, recording the resulting bound on $H^{\epsilon_s}_{\min}$. To ensure that we approach the asymptotic bound for increasing n, we set $\gamma = \delta_1 = \cdots = \delta_{|\mathcal{G}|} = n^{-1/3}$. This choice also forces the completeness error to be constant across all values of n.

to increase the certifiable entropy by considering fewer scores. However, it is still worth noting that this entropy loss could be reduced by choosing a more generous set of protocol parameters, e.g., increasing n and decreasing δ.

In practice, increasing n can be difficult due to restrictions on the overall runtime of the protocol. Not only does it take longer to collect the statistics within the device-interaction phase, but it may also increase the runtime of the extraction phase [98]. In Fig. 4.3 we observe how quickly the various protocols converge on their respective asymptotic rates as we increase n. Again we find that, due to finite-size effects, entropy loss is far greater for $\mathcal{G}^{23}_{\mathrm{EB}}$ than for the other protocols. In particular, we see that for protocols with fewer than 10^{10} rounds, it is advantageous to use $\mathcal{G}^{23}_{\langle AB\rangle}$. From the perspective of practical implementation, Fig. 4.2c and Fig. 4.3 highlight the benefits of a flexible protocol framework. Looking at the results, there is no best protocol for all scenarios. Rather, in order to maximise the quantity of randomness gained, the user should utilise the flexible construction and design their own protocol tailored to the scenario under consideration.

It is also important to compare the rates of instances of Protocol QRE with other protocols from the literature, in particular the protocol of [26] (ARV). In [26], the min-tradeoff

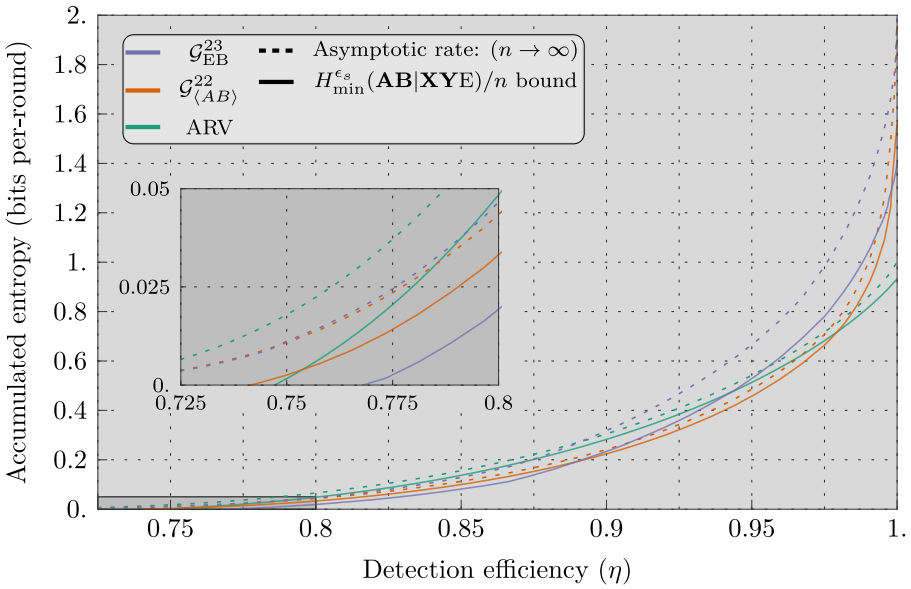


Figure 4.4: Comparison between the certifiable accumulation rates of QRNE protocols based on $\mathcal{G}_{\mathrm{CHSH}}$, $\mathcal{G}^{23}_{\mathrm{EB}}$ and Protocol ARV from [26] on qubit systems with inefficient detectors (cf. Fig. 4.2). The rates of Protocol ARV are also evaluated using the improved EAT statement [25]. For Protocol ARV, we use the one-sided von Neumann entropy bound, so the maximum rate is one bit per round, but because we can directly get the single-round von Neumann entropy, the rate initially falls more slowly with decreasing detection efficiency than for the other protocols.

functions are constructed from a tight bound on the single-party von Neumann entropy, $H(A|XE)$, which is given in terms of a CHSH inequality violation [72]. In Fig. 4.4 we compare the rates of ARV with $\mathcal{G}^{22}_{\langle AB\rangle}$ and $\mathcal{G}^{23}_{\mathrm{EB}}$ for entangled qubit systems with inefficient detectors. To make our comparison fair, we have also computed the rates for Protocol ARV using the improved EAT bound[8]. As the rates of Protocol ARV are derived from the entropy accumulated by a single party their rates are capped at one bit per round.

In contrast, the semidefinite programs grant us access to bounds on the entropy produced by both parties and we are therefore able to certify up to two bits per round. In Fig. 4.4, this advantage is observed in the high detection efficiency regime. Fig. 4.4 also highlights a significant drawback of our technique, which stems from our use of the inequality $H(AB|XYE) \geq H_{\min}(AB|XYE)$. In particular, we see that for $\eta < 0.9$, the $H(A|XE)$ bound for the CHSH inequality is already greater than the $H_{\min}(AB|XYE)$ established for the empirical behaviour. Therefore, in the asymptotic limit $(n \to \infty)$ the min-entropy bounds for these protocols will produce strictly worse rates in this regime. For the finite n we have chosen, $n = 10^{10}$, it appears that for the majority of smaller η, it is advantageous to use the

[8]Note that we always use the direct bound on the von Neumann entropy when considering Protocol ARV, rather than forming a bound via the min-entropy

ARV protocol over the protocols derived from the framework. Nevertheless, looking at the threshold detection efficiencies, i.e. the minimal detection efficiency required to achieve positive rates, we find that some protocols from our framework are able to again beat the rates established for Protocol ARV. Looking at the inset plot in Fig. 4.4 we see that $\mathcal{G}^{22}_{\langle AB\rangle}$ has a smaller threshold efficiency than that of Protocol ARV for the chosen protocol parameters. Interestingly, this shows that $\mathcal{G}^{22}_{\langle AB\rangle}$ is capable of producing higher rates than Protocol ARV in both the low and the high detection efficiency regimes, with the improvement for low detection efficiencies being of particular relevance to experimental implementations. Importantly, this shows that protocols from the framework are of practical use for finite n in spite of the losses coming from the use of $H(AB|XYE) \geq H_{\min}(AB|XYE)$.

4.3 Net-positive expansion rates with current technologies

In recent years, a few experiments have reported implementations of randomness expansion protocols [89, 99] and the successful device-independent certification of randomness. However, none of these experiments managed to fully account for the cost of seeding their protocol and so a net-gain in entropy (or 'expansion') is yet to be seen.[9] Here we investigate the question as to whether expansion is possible using current technologies. By optimising the protocol of [26], we find that net-positive rates are indeed within the capabilities of current nonlocality experiments.

For this analysis we choose to move away from our framework and adapt the protocol of Arnon-Friedman, Renner and Vidick [26] (Protocol ARV). In the security proof of Protocol ARV a direct analytical bound on the von Neumann entropy for a single agent's outputs is established using the results of [72].

4.3.1 Protocol ARV

For clarity let us very briefly review the Protocol introduced in [26]. The protocol follows the same spot-checking structure presented in Chapter 3, with both agents having binary inputs and outputs. On test rounds the agents play the CHSH game and their respective generation inputs are $(\tilde{x},\tilde{y}) = (0,0)$. We denote by $\omega \in [\frac{3}{4}, \frac{1}{2} + \frac{\sqrt{2}}{4}]$ the agents' expected score for the CHSH game. By Equation (4.7) in [26] we can lower bound the von Neumann entropy of a single interaction as

$$\inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i^{\text{test}}(\sigma)_{C_i} = \tau_\omega} H(A_iB_i|X_iY_iR')_{\mathcal{N}_i(\sigma)} \geq 1 - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega(\omega-1)+3}\right) \tag{4.3.1}$$

where $\tau_\omega = (1-\omega)|0\rangle\langle 0| + \omega|1\rangle\langle 1|$.

[9] In [99] the authors argue that costs of seeding the protocol are moot if one uses a public source of randomness such as the NIST randomness beacon [100]. Whilst this may be true, one is then limited by the speed of the public source. For the case of the NIST beacon this seed generation rate is 512 bits every 60 seconds. At this speed, for the protocol in Example 3.3 the accumulation stage would need to run for around 10^6 minutes or just under 2 years!

In [26] this bound is used in conjunction with the original statement of the EAT [24]. To further improve upon the rates we use the improved EAT bound [25]. We note that the adaptation of Protocol ARV to the improved EAT bound was already constructed in [25]. Our adaptation follows roughly the same procedure, with a few additional modifications to further increase its practicality in the high noise regime.

Min-tradeoff function construction

Let $g(\omega) := 1 - h\left(\frac{1}{2} + \frac{1}{2}\sqrt{16\omega(\omega-1)+3}\right)$, we can construct affine, entropy-bounding functions from g by taking linear approximations at some point $\nu \in \left(\frac{3}{4}, \frac{1}{2} + \frac{\sqrt{2}}{4}\right)$. That is, we define

$$g_\nu(\omega) := g(\nu) + g'(\nu)(\omega - \nu), \tag{4.3.2}$$

with $g' = \frac{\mathrm{d}g}{\mathrm{d}\omega}$. As g is a convex function, g_ν will also lower bound the von Neumann entropy $H(A_i B_i | X_i Y_i R')$. Writing this in the language used within Chapter 3, we define the constant $\alpha_\nu = g(\nu) - \nu g'(\nu)$ and then

$$g_\nu(\boldsymbol{\delta}_c) = \begin{cases} \alpha_\nu + g'(\nu) & \text{for } c = 1 \\ \alpha_\nu & \text{for } c = 0 \end{cases}. \tag{4.3.3}$$

As $g'(\nu) > 0$ for all $\nu \in (\frac{3}{4}, \frac{1}{2} + \frac{\sqrt{2}}{4})$ we immediately have $\mathrm{Max}[g_\nu] = \alpha_\nu + g'(\nu)$.

Instead of applying Lemma 3.1 to g_ν in order to construct a min-tradeoff function, we consider a more general extension. Let $\alpha, \lambda_0, \lambda_1, \lambda_\perp \in \mathbb{R}$ and define

$$f_\nu(\boldsymbol{\delta}_c) = \begin{cases} \alpha + \lambda_1 & \text{for } c = 1 \\ \alpha + \lambda_0 & \text{for } c = 0 \\ \alpha + \lambda_\perp & \text{for } c = \perp \end{cases}. \tag{4.3.4}$$

The function f_ν is an arbitrary affine function on distributions over $\mathcal{C} = \{1, 0, \perp\}$. For f_ν to also be a min-tradeoff function we require that if it is evaluated for any protocol respecting distributions then it should bound the von Neumann entropy (cf. (2.3.6)). A sufficient condition for this is that f_ν and g_ν are equal when evaluated at $(\gamma\omega, \gamma(1-\omega), 1-\gamma)$ and $(\omega, 1-\omega)$ respectively. More explicitly, we require that for $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{\sqrt{2}}{4}]$,

$$\alpha_\upsilon + \omega g'(\nu) = \alpha + \gamma\omega\lambda_1 + \gamma(1-\omega)\lambda_0 + (1-\gamma)\lambda_\perp. \tag{4.3.5}$$

As this must hold for all ω simultaneously, we may split this into two conditions:

$$\alpha_\nu = \alpha + \gamma\lambda_0 + (1-\gamma)\lambda_\perp \tag{4.3.6}$$

and

$$g'(\nu) = \gamma\lambda_1 - \gamma\lambda_0. \tag{4.3.7}$$

Rearranging (4.3.6) we find that

$$\lambda_0 = \frac{\alpha_\nu - \alpha - (1-\gamma)\lambda_\perp}{\gamma}. \tag{4.3.8}$$

We may now also solve for λ_1 in (4.3.7),

$$\lambda_1 = \frac{g'(v)}{\gamma} + \frac{\alpha_v - \alpha - (1-\gamma)\lambda_\perp}{\gamma}. \tag{4.3.9}$$

Therefore, we may parametrize our extension by the pair $(\alpha, \lambda_\perp)$, i.e.

$$f_v(\boldsymbol{\delta}_c) = \begin{cases} \alpha + \frac{g'(v)}{\gamma} + \frac{\alpha_v - \alpha - (1-\gamma)\lambda_\perp}{\gamma} & \text{for } c = 1 \\ \alpha + \frac{\alpha_v - \alpha - (1-\gamma)\lambda_\perp}{\gamma} & \text{for } c = 0 \\ \alpha + \lambda_\perp & \text{for } c = \perp \end{cases}. \tag{4.3.10}$$

We can then optimize our choice of min-tradeoff function not only over v but also over the parameters α and $\lambda_\perp$.

Remark 4.2. The extension presented in Lemma 3.1 is equivalent to choosing $\alpha = \alpha_v$ and $\lambda_\perp = g'(v)$.

4.3.2 A sharper completeness error

For $\delta > 0$ and an expected CHSH score ω, the event that protocol ARV does not abort after n rounds is

$$\Omega_{ARV} = \{C_1^n \mid \gamma(\omega - \delta) < F_{C_1^n}(1)\}. \tag{4.3.11}$$

In order to minimize the required runtime of the experiment we employ a stronger concentration bound to the task of computing the completeness error than that which was used for the framework. The following lemma from [101] gives a good approximation to the cumulative distribution function of a binomial process.

Lemma 4.1. *Let $X \sim \mathrm{Bin}(n,p)$, then for every $k = 0, 1, \ldots, n-1$ and every $p \in (0,1)$, we have*

$$C_{n,p}(k) \leq \mathbb{P}[X \leq k] \leq C_{n,p}(k+1) \tag{4.3.12}$$

where

$$C_{n,p}(k) := \Phi\left(\mathrm{sgn}\left(\tfrac{k}{n} - p\right)\sqrt{2nD(k/n||p)}\right), \tag{4.3.13}$$

with $D(x||p) := x\ln(x/p) + (1-x)\ln((1-x)/(1-p))$ and $\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$.

Corollary 4.1 (Completeness error)**.** *For $0 < \delta < \omega$, Protocol ARV has a completeness error*

$$\varepsilon_{comp} = C_{n,\gamma\omega}\left(\lceil \gamma(\omega - \delta)n \rceil + 1\right). \tag{4.3.14}$$

Proof. If an implementation of Protocol ARV is honest then $F_{C_1^n}(1) \sim \frac{1}{n}\mathrm{Bin}\left(n, \gamma\omega\right)$. So, the probability the protocol aborts can be bounded above as

$$\begin{aligned} \mathbb{P}\left[\Omega^c\right] &= \mathbb{P}\left[F_{C_1^n}(1) \leq \gamma(\omega - \delta)\right] \\ &= \mathbb{P}\left[nF_{C_1^n} \leq \gamma(\omega - \delta)n\right] \\ &\leq \mathbb{P}\left[nF_{C_1^n} \leq \lceil \gamma(\omega - \delta)n \rceil\right] \\ &\leq C_{n,\gamma\omega}\left(\lceil \gamma(\omega - \delta)n \rceil + 1\right). \end{aligned} \tag{4.3.15}$$

■

4.3.3 Application to realistic parameter regimes

Through correspondence with the experimental group at the University of Science and Technology of China [102] we gathered estimates for the current capabilities of photonics based nonlocality experiments. They report a laser pulse of 2MHz (2×10^6 rounds per second) with an expected CHSH score of 0.75132. Restricting the completeness error to be no larger than 10^{-3} and $\epsilon_s = \epsilon_{\text{EAT}} = 10^{-4}$, we used a numerical search to find a set of parameters (n, γ, δ) which gave a positive net gain in entropy and, at the same time, minimized n. This search was fairly rudimentary, we defined a lattice of points for (n, γ, δ) and proceeded to calculate the net gain in entropy and the completeness error at each point. We then took the point with the smallest n that had a completeness error less that 10^{-3} and which produced a positive net-rate. This numerical search found that the following parameter choices,

Quantity	Value
n	9.55×10^{11}
γ	1.50×10^{-4}
δ	2.24×10^{-4}
$\varepsilon_{\text{comp}}$	9.86×10^{-4}

achieved a net-gain in entropy of 1.93×10^{-5} bits per round, with a total accumulation procedure time of around 5 days. Which is within the bounds of current nonlocality experiments – longer experiments would begin to suffer from stability issues and impractical runtimes.

Fixing γ and δ, we can vary n to illustrate how the various improvements pushed the rates into experimentally achievable regimes. In Fig. 4.5 we compare the two separate derivations of the completeness errors. Observing where the two curves drop below a completeness error of 10^{-3}, we see that using the improved bound allows us to half the overall number of rounds in our experiment. Similarly, in Fig. 4.6 we compare the two separate statements of the EAT, [25] and [24]. By using the improved statement of the entropy accumulation theorem we are able to achieve a net-gain in randomness using a whole order of a magnitude fewer rounds.

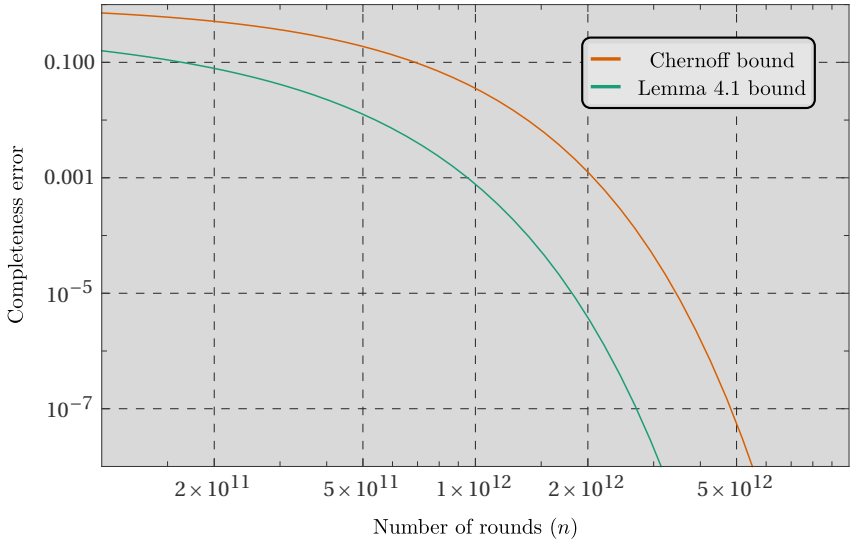


Figure 4.5: A comparison between the completeness error derived from Lemma 4.1 and the completeness error derived from the Chernoff bound.

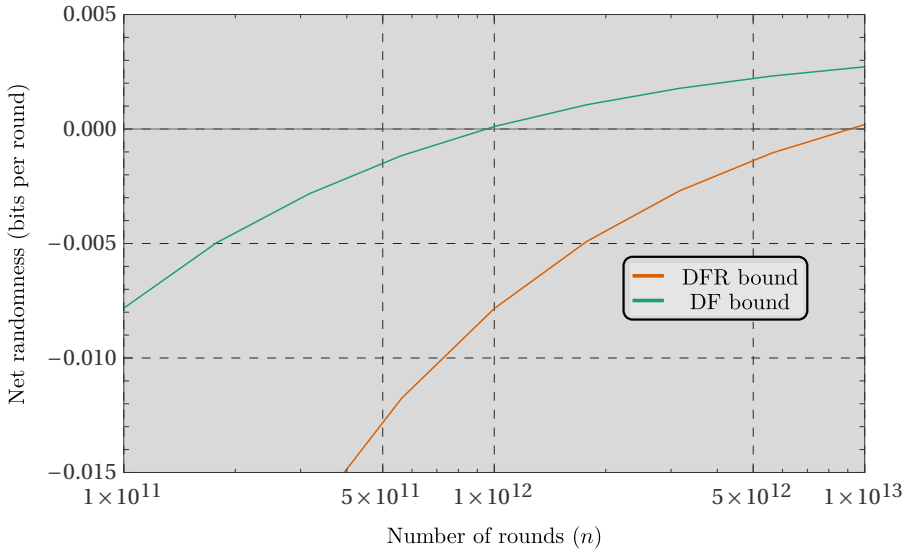


Figure 4.6: A comparison of the net gain/loss in randomness certified when using the two different statements of the EAT. The green line (DF bound) refers to the net randomness rate certified using the improved EAT statement [25] whereas the orange line (DFR bound) refers to the bound established in the appendix of [26], which relies on the original statement of the EAT [24].

CHAPTER 5

CONCLUSIONS AND OUTLOOK

This thesis sought progress towards a more practical future for device-independent randomness expansion. In Chapter 3, we introduced a framework for building quantum-secure randomness expansion protocols. By combining device-independent bounds on the guessing probability with the EAT, we were able to achieve full quantum security for spot-checking protocols based on any nonlocality tests. Moreover, through semidefinite programming techniques this procedure can be made computationally efficient. A key advantage of this approach is that it allows a user to freely modify their choice of nonlocality test in order to better accommodate the scenario within which they are generating randomness. This is especially useful within the context of device-independence as we cannot assume an ability to tune our untrusted devices to better fit pre-existing protocols. However, through this flexible protocol construction, we may now tune our protocol to better fit the devices.

In Chapter 4 we introduced examples of protocols built within the framework. We modelled their implementation on entangled qubit systems and analysed their robustness to inefficient detectors – a significant source of noise in photonics-based implementations. Our analysis showed a tradeoff between the complexity of the chosen nonlocality test and our confidence in the resulting statistics collected over a finite number of trials. Whilst the more complex nonlocality tests provide a stronger characterisation of the untrusted devices, we found that their requirement for a large number of trials led them to be outperformed by simpler tests when used within protocols with smaller numbers of rounds. Further reinforcing the need for a user to be able to adapt the protocol to fit their available resources. We also compared the rates of a selection of our protocols to the protocol presented in [26] (ARV). Interestingly, we found that some of the protocols from the framework are able to achieve higher rates than Protocol ARV in both the high and low detection efficiency regimes. In particular, the higher rates for low detection efficiencies is of great importance

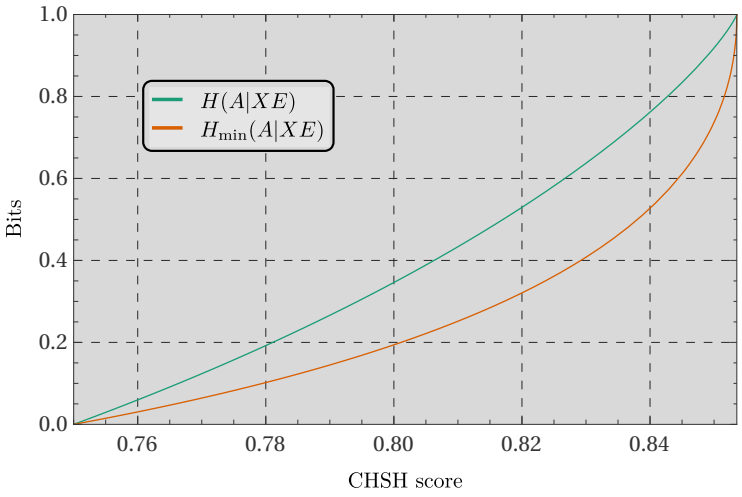


Figure 5.1: A plot of tight lower bounds on $H(A|XE)$ and $H_{\min}(A|XE)$ in terms of the CHSH score. The bound on $H(A|XE)$ is given by (4.3.1) and the bound on $H_{\min}(A|XE)$ is taken from [18].

for actual experimental implementations.

Our analysis also led us to investigate the performance of current randomness expansion experiments. Previous experiments [89,99] have reported successful device-independent certification of randomness. However, no experiment has yet to demonstrate full randomness expansion, i.e. a net increase in total entropy. Through careful application of the improved EAT statement [25] to the protocol of [26] we were able to find a set of protocol parameters that could achieve full randomness expansion which were, crucially, within the realms of current experimental capabilities [102].

Although the framework produces secure and robust protocols, there remains scope for further improvements. For example, optimising the choice of min-tradeoff function is a non-convex and not necessarily continuous problem [103]. Our analysis in Sec. 4.1 used a simple probabilistic gradient ascent algorithm to approach this problem. A more sophisticated approach to this optimization could yield higher EAT-rates, particularly for protocols with a higher number of scores, e.g., those which used the full behaviour as their nonlocality test ($\mathcal{G}_{\text{EB}}$).

The construction of min-tradeoff functions using SDPs was made possible by the relation $H(AB|XYE) \geq H_{\min}(AB|XYE)$. Unfortunately, this inequality is not generally tight. One can observe this difference in the context of the CHSH game, where $H(AB|XYE)$ and $H_{\min}(AB|XYE)$ both admit an analytical form (see Fig. 5.1).[1] Several alternative approaches could be taken in order to reduce this loss. Firstly, the von Neumann entropy

[1] More recently, the authors of [104] showed that for this particular scenario one can also derive a tight bound on the Rényi entropy of order 2, $H_2(A|XE)$. Moreover, their bound coincides with the bound for $H_{\min}(A|XE)$.

and the min-entropy are special cases of a larger family of entropies, the Rényi entropies $H_\alpha(AB|XYE)$ [50] and the relation $H(AB|XYE) \geq H_{\min}(AB|XYE)$ is part of a more general ordering on this family,[2] So, if we are able to develop efficient computational techniques for computing device-independent lower bounds on some of these other entropies (those lying between $H(AB|XYE)$ and $H_{\min}(AB|XYE)$) then this would likely lead to an immediate improvement on the rates of certifiable randomness. In general, these quantities are more difficult to evaluate than $H_{\min}$ as they are nonlinear expressions of the quantum state. One possible approach to this problem would be to introduce a density operator, which acts as the quantum state shared between the untrusted devices, into the set of operators considered within the NPA hierarchy. This would permit relaxations of expressions that are nonlinear in the state, which could allow one to approximate H_α for fractional values of α. Alternatively, one could directly introduce state nonlinear terms by including operators akin to $\mathbb{1}\langle M_{a|x}\rangle$. This was recently proposed in [105] as a method of approximating correlations emerging from quantum networks.

In certain scenarios, dimension-dependent bounds may also be applicable. For example, it is known that for the special case of n-party, 2-input, 2-output scenarios it is sufficient to consider qubit systems [72]. It may therefore be possible to derive results, analogous to those of [72], for nonlocality tests such as the GHZ game [106]. Furthermore, with a dimension bound one may be able to adapt the recent numerical techniques of [107] which give robust lower bounds on the von Neumann entropy for device-dependent protocols.

As the framework permits protocols that rely on any nonlocality test, it is natural to then search for tests that provide high randomness certification rates. Investigations into the randomness certification properties of nonlocality tests with larger output alphabets or additional parties could be of interest. However, increasing either of these parameters is likely to increase the influence of finite-size effects. Alternatively, one could try to design more economical nonlocality tests by combining scores that are of a lesser importance to the task of certifying randomness. Intuitively, for a score $c \in \mathcal{C}$, the magnitude of of $\lambda(c)$ in the min-tradeoff function indicates how important that score is for certifying entropy. If $|\lambda(c)|$ is large then this score is 'important' in the sense that any small deviations in the expected frequency of that score, $\omega(c)$, will have a large impact on the amount of certifiable entropy. Another approach to designing good nonlocality tests would be to take inspiration from [37, 38] wherein the authors showed how to derive the optimal Bell-expressions for certifying randomness. A nonlocal game could then be designed to encode the constraints imposed by this optimal Bell-expression. An example of such a game would be to assign a score $+1$ to all $(ABXY)$ that have a positive coefficient in the optimal Bell-expression and a score of -1 to all those with negative coefficients. The input distribution of the nonlocal game could then be chosen as such to encode the relative weights of the coefficients.

[2] The Rényi entropies are one of many different entropic families that include the von Neumann entropy as a limiting case. Any such family could be used if they satisfy an equivalent relation.

Finally, our computational approach to the EAT considered only the task of randomness expansion. Our work could be extended to produce adaptable security proofs for other device-independent protocols. Given that the EAT has already been successfully applied to a wide range of problems [29–31, 73, 108], developing good methods for robust min-tradeoff function constructions represents an important step towards practical device-independent security.

Protocol notation

Notation	Description
$\mathfrak{D}$	A collection of untrusted devices.
$\mathcal{G}$	A nonlocal game.
$\Omega_{\mathcal{G}}$	Set of expected distributions over scores of $\mathcal{G}$ using quantum strategies.
$\Omega_{\mathcal{G}}^{(k)}$	Set of expected distributions over scores of $\mathcal{G}$ using strategies from $\Omega^{(k)}$.
$\boldsymbol{\nu}, \boldsymbol{\omega}$	Expected distributions over scores of some nonlocal game $\mathcal{G}$.
$\boldsymbol{\lambda}_{\boldsymbol{\nu}}$	Dual feasible point of the guessing probability program parametrized by $\boldsymbol{\nu}$.
$\boldsymbol{\delta}$	Vector of statistical confidence interval widths.
$\boldsymbol{\delta}_{\pm}$	$\boldsymbol{\delta}$ with elements signed in accordance with a given $\boldsymbol{\lambda}_{\boldsymbol{\nu}}$.
$\mathcal{A}, \mathcal{B}$	Devices' output alphabets.
$\mathcal{X}, \mathcal{Y}$	Devices' input alphabets.
$n \in \mathbb{N}$	Number of rounds in the device-interaction phase.
$\gamma \in (0,1)$	Probability that any given round is a test round.
A_i, B_i	Devices' outputs for the i^{th} round.
X_i, Y_i	Devices' inputs for the i^{th} round.
C_i	Score recorded for the i^{th} round.
Ω	Event that the protocol does not abort.
$\boldsymbol{F}_{C_1^n}$	Empirical frequency distribution arising from C_1^n.
$\varepsilon_{\text{comp}}$	Completeness error of Protocol QRE.
$\varepsilon_{\text{sound}}$	Soundness error of Protocol QRE.
ϵ_s	Smoothing parameter for $H_{\min}$.
ϵ_{EAT}	Tolerance of unlikely success events.
ϵ_V	EAT error term (Variance).
ϵ_K	EAT error term (Remainder).
ϵ_Ω	EAT error term (Success probability).
R_{ext}	Strong quantum-secure randomness extractor.
ϵ_{ext}	Extractor error.
ℓ_{ext}	Entropy lost during extraction.

APPENDIX A

FINITE PRECISION SECURITY

The framework of Chapter 3 bases its security on a numerical computation which, in practice, is at the mercy of finite precision computing. Rounding errors will be present within the computations and in the worst case they could surreptitiously collude to overestimate the accumulated entropy, falsifying any subsequent security statements. In this section we will show how to account for these errors, making the min-tradeoff functions robust to finite precision computation.

Recall the generic semidefinite program

$$\begin{aligned} \sup_{X \in \mathfrak{S}_n(\mathbb{R}^n)} \quad & \mathrm{Tr}[CX], \\ \text{subject to.} \quad & \mathrm{Tr}[F_i X] = b_i \quad \text{for } i \in 1, \ldots, r, \\ & X \succeq 0, \end{aligned} \tag{A.0.1}$$

and its dual form

$$\begin{aligned} \inf_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & \boldsymbol{\lambda}^T \cdot \boldsymbol{b} \\ \text{subject to.} \quad & C - \sum_i \lambda_i F_i \preceq 0 \end{aligned} \tag{A.0.2}$$

We would like to know how well the dual functional $\boldsymbol{b}' \mapsto \boldsymbol{\lambda} \cdot \boldsymbol{b}'$ preserves its upper-bounding property (cf. (2.7.8)) when we allow for small perturbations in the constraints. To model this we introduce an error parameter $\delta \geq 0$. We then rewrite the dual program as

$$\begin{aligned} \inf_{\boldsymbol{\lambda} \in \mathbb{R}^m} \quad & \boldsymbol{\lambda}^T \cdot \boldsymbol{b} \\ \text{subject to.} \quad & C - \sum_i \lambda_i F_i \preceq \delta \mathbb{1} \end{aligned} \tag{A.0.3}$$

Now let $\hat{\boldsymbol{b}}$ be another constraint vector, let $\hat{X}$ be any feasible point of the primal program constrained by $\hat{\boldsymbol{b}}$ and let $\hat{p} = \mathrm{Tr}\left[C\hat{X}\right]$. In the error free dual program (A.0.2), we can apply

the map $M \mapsto \mathrm{Tr}\left[\hat{X}M\right]$ to the linear matrix inequality constraint to find that

$$\hat{p} \leq \boldsymbol{\lambda} \cdot \hat{\boldsymbol{b}}, \tag{A.0.4}$$

i.e. the upper bounding property of the dual functional. Applying the same map to the perturbed dual (A.0.3) we get

$$\hat{p} \leq \boldsymbol{\lambda} \cdot \hat{\boldsymbol{b}} + \delta \,\mathrm{Tr}\left[\hat{X}\right]. \tag{A.0.5}$$

Therefore, the amount to which the dual functional can fail to satisfy the upper bounding property (A.0.4) is exactly $\delta \,\mathrm{Tr}\left[\hat{X}\right]$. We now explain how to bound δ and $\mathrm{Tr}\left[\hat{X}\right]$.

Since most semidefinite programming solvers will report back the degree to which the various constraints have been satisfied, the value of δ can be readily observed. A problem is considered feasible in practice, if each of its constraints have been satisfied up to some small perturbation. Roughly, an off-the-shelf SDP solver would consider a solution feasible if δ is no greater than $10^{-5} \sim 10^{-8}$. By using high precision solvers, e.g. those of the sdpa family [84], one can further reduce the value of δ.

To bound the expression $\mathrm{Tr}\left[\hat{X}\right]$ it helps to consider the actual SDP of interest, i.e. the guessing probability program (2.7.13) relaxed to the k^{th} level of the NPA hierarchy. In this context, each element of $\hat{X}$ represents some expression of the form $\langle\psi| M_1 M_2 \ldots M_n |\psi\rangle$, where $|\psi\rangle \in \mathcal{H}$ and $M_1, \ldots, M_n$ are bounded, self-adjoint operators on some Hilbert space $\mathcal{H}$. For $M \in \mathcal{L}(\mathcal{H})$, the operator norm of M is $\|M\| = \sup_{|\psi\rangle\in\mathcal{H}}\{\|M|\psi\rangle\| \mid \||\psi\rangle\| \leq 1\}$, if in addition M is self-adjoint then we also have $\|M\| = \sup_{|\psi\rangle\in\mathcal{H}}\{|\langle\psi|M|\psi\rangle| : \||\psi\rangle\| \leq 1\}$ (see Proposition 2.2 in [109]). Note that the operator norm is also sub-multiplicative, i.e. for $M, N \in \mathcal{L}(\mathcal{H})$ we have $\|MN\| \leq \|M\|\|N\|$. Therefore, for the element $\langle\psi| M_1 M_2 \ldots M_n |\psi\rangle$ of $\hat{X}$ we have

$$\begin{aligned} |\langle\psi| M_1 M_2 \ldots M_n |\psi\rangle| &\leq \||\psi\rangle\| \|M_1 M_2 \ldots M_n |\psi\rangle\| \\ &\leq \|M_1 M_2 \ldots M_n\| \\ &\leq \|M_1\|\|M_2\| \ldots \|M_n\| \end{aligned} \tag{A.0.6}$$

where the first line follows from the Cauchy-Schwarz inequality $|\langle\psi|\phi\rangle| \leq \||\psi\rangle\| \||\phi\rangle\|$. Furthermore, as each M_i is POVM element, we have $0 \leq \langle\psi|M_i|\psi\rangle \leq 1$ and so $\|M_i\| \leq 1$. It follows that every element of $\hat{X}$ is bounded by 1. If $\hat{X}$ is a $d \times d$ matrix, we have $\mathrm{Tr}\left[\hat{X}\right] \leq d$.

It remains to bound the size of $\hat{X}$. Recall that for the k^{th} level of the hierarchy, $\hat{X}$ is indexed by all unique products of operators of length no larger than k. The generating set of operators forms the 1^{st} level of the hierarchy, in our case this is $\mathcal{W}^{(1)} = \{\mathbb{1}\} \cup \{E_{a|x}\}_{(a,x)\in\mathcal{A}\mathcal{X}} \cup \{F_{b|y}\}_{(b,y)\in\mathcal{B}\mathcal{Y}}$, i.e. the POVM elements that form our distribution $p(a,b|x,y) = \mathrm{Tr}\left[\rho(E_{a|x}F_{b|y})\right]$. In fact, we are at liberty to use a slightly smaller set: after choosing some $(a',b') \in \mathcal{A}\mathcal{B}$, let $\mathcal{A}' = \mathcal{A} \setminus \{a'\}$ and $\mathcal{B}' = \mathcal{B} \setminus \{b'\}$ and define

$$\tilde{\mathcal{W}}^{(1)} = \{\mathbb{1}\} \cup \{E_{a|x}\}_{(a,x)\in\mathcal{A}'\mathcal{X}} \cup \{F_{b|y}\}_{(b,y)\in\mathcal{B}'\mathcal{Y}}.$$

Where we have removed the operators $\{E_{a'|x}\}_{x\in\mathcal{X}} \cup \{F_{b'|y}\}_{y\in\mathcal{Y}}$ as they can be constructed from linear combinations of the remaining operators, i.e. $E_{a'|x} = \mathbb{1} - \sum_{a\in\mathcal{A}'} E_{a|x}$. This reduction of

the generating set is a consequence of Lemma 7 in [41], which states that if a certificate exists for a word set $\mathcal{W}$ then a certificate will also exist for a word set $\tilde{\mathcal{W}}$ where each word in $\tilde{\mathcal{W}}$ is a linear combination of words from $\mathcal{W}$. By decreasing the size of $\hat{X}$ we speed up the computation and reduce the impact of floating point errors.

For the 1st level of the hierarchy, $\hat{X}$ is a $d^{(1)} \times d^{(1)}$ matrix with

$$d^{(1)} = 1 + (|\mathcal{A}| - 1)|\mathcal{X}| + (|\mathcal{B}| - 1)|\mathcal{Y}|.$$

The second level of the hierarchy is indexed by all words present at the first level together with all unique, non-trivial products of operators of length 2. There are two sources of redundancy here: we assume that our measurements are projective and the measurement operators for separate parties commute. The projective assumption means that when we only consider a single party, the only nontrivial products are those formed from operators that correspond to different inputs. Therefore, at the second level of the hierarchy there are $(|\mathcal{A}| - 1)|\mathcal{X}| \times (|\mathcal{A}| - 1)(|\mathcal{X}| - 1)$ new words of the form $E_{a|x}E_{a'|x'}$; $(|\mathcal{B}| - 1)|\mathcal{Y}| \times (|\mathcal{B}| - 1)(|\mathcal{Y}| - 1)$ new words of the form $F_{b|y}F_{b'|y'}$ and $(|\mathcal{A}| - 1)(|\mathcal{B}| - 1)|\mathcal{X}||\mathcal{Y}|$ words of the form $E_{a|x}F_{b|y}$. Meaning that,

$$\begin{aligned} d^{(2)} = {} & 1 + (|\mathcal{A}| - 1)|\mathcal{X}| + (|\mathcal{B}| - 1)|\mathcal{Y}| \\ & + (|\mathcal{A}| - 1)^2|\mathcal{X}|(|\mathcal{X}| - 1) + (|\mathcal{B}| - 1)^2|\mathcal{Y}|(|\mathcal{Y}| - 1) + (|\mathcal{A}| - 1)(|\mathcal{B}| - 1)|\mathcal{X}||\mathcal{Y}|. \end{aligned}$$

In general $d^{(k)}$ will grow exponentially with the level of the hierarchy. This means that for higher levels of the hierarchy we may have to use a higher precision solver to curb non-negligible error propagation.[1] Fortunately, all of our computations were performed at the 2nd level of the hierarchy and so the product $\delta \operatorname{Tr}\left[\hat{X}\right]$ was always small for the size of input and output alphabets.

Remark A.1. Actually, for the guessing probability program (2.7.13), the certificate $\hat{X}$ is composed of $|\mathcal{AB}|$ blocks, i.e. $\hat{X} = \bigoplus_{ab \in \mathcal{AB}} \hat{X}_{ab}$ – one block for each subnormalized distribution $\tilde{\boldsymbol{p}}_{ab} \in \mathfrak{Q}^{(k)}$, with the block $\hat{X}_{ab}$ corresponding to a NPA certificate for the distribution $\tilde{\boldsymbol{p}}_{ab}$. To account for the subnormalization of $\tilde{\boldsymbol{p}}_{ab}$, each block is weighted by the norm $|\tilde{\boldsymbol{p}}_{ab}| = \sum_{a'b'} \tilde{p}_{ab}(a', b'|x, y)$. However, as $\operatorname{Tr}\left[\hat{X}\right] = \sum_{ab} |\tilde{\boldsymbol{p}}_{ab}| \operatorname{Tr}\left[\hat{X}_{ab}\right]$, we can upper bound $\operatorname{Tr}\left[\hat{X}\right]$ by just considering a single (normalized) block.

[1] Larger matrices and higher precision solvers will have a compounding negative effect on the speed of computation.

APPENDIX B

AN IMPLEMENTATION OF THE FRAMEWORK IN PYTHON

To facilitate the framework's use, a python implementation was developed and released [42]. The package provides a user-friendly means of designing and computing the relevant security quantities of framework based protocols. We now provide a short overview of the package including its basic structure and functionality.

The package is built around three core objects: *games*, *devices* and *protocols*.

Games

The *game* object allows a user to specify a real-valued nonlocal game, an expected score for this game and a statistical confidence in that expected score. The scoring rule $V : \mathcal{ABXY} \to \mathbb{R}$ is specified by supplying a matrix of coefficients following the same indexing pattern present in Fig. 4.1. E.g., for the CHSH game (cf. Ex. 2.3) this would be

$$\begin{pmatrix} 0.25 & 0 & 0.25 & 0 \\ 0 & 0.25 & 0 & 0.25 \\ 0.25 & 0 & 0 & 0.25 \\ 0 & 0.25 & 0.25 & 0 \end{pmatrix}. \tag{B.0.1}$$

To implement games with multiple scores one would create a collection of game objects, one for each score. For example, the game $\mathcal{G}_{\langle AB \rangle}$ for binary $\mathcal{ABXY}$ can be implemented with 4

game objects for each of its separate scores whose coefficient matrices are

$$\begin{pmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0.25 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0.25 \end{pmatrix}.$$

As we have weighted the coefficients with the input distribution (assumed uniform), the expected scores for these individual games refer to the relative frequency of the score they represent. The normalization score will be handled automatically by the program.

Devices

The *devices* object is initialized by supplying a list of attributes including: the input-output configuration, the generation inputs, any nonlocal games played by the device and a desired SDP relaxation level. Once initialized, the code uses another package 'ncpol2sdpa' [110] to create a relaxation of the guessing probability program corresponding to the specified attributes. A user may then request the guessing probability or min-entropy generated when the generation inputs are used, the program then solves the SDP and returns the relevant value. A user is free to alter the devices' attributes post-initialization, upon doing so the program automatically updates the guessing probability program and any subsequent requests for the min-entropy will take into account these changes.

Protocol

The *protocol* object stores all of the additional protocol parameters required by Protocol QRE including: number of rounds, testing probability, ϵ_s and ϵ_{EAT}.

Usage

Once the protocol and device objects have been initialized, a user can begin to compute quantities relevant to Protocol QRE. In particular, the completeness error and all of the relevant EAT quantities (3.2.7) – (3.2.10) may be computed for a specified min-tradeoff function from $\mathcal{F}_{\min}$. The default min-tradeoff function is the one indexed by the expected scores provided. The code also implements a rudimentary gradient ascent algorithm to optimize the choice of min-tradeoff function, this was used in Ex. 3.2 and Fig. 3.2.

APPENDIX C

BLOCKING THE SPOT-CHECKING PROTOCOL

The original statement of the entropy accumulation theorem [24] was released alongside an accompanying work, [26], which detailed its application to security proofs of device-independent protocols. Within the appendix of [26] it was shown that one could increase the quantity of entropy certified by the original EAT through a modification to the structure of the spot-checking protocol. In light of this the authors of [25] improved the second order term of the EAT in order to account for the suboptimal dependence on the testing probability that was highlighted by the structural modification. In the sections that follow, we show how the family of min-tradeoff functions $\mathcal{F}_{\min}$ can be adapted to this structural change. Furthermore, we compare the accumulation rates achievable with the different structures and EAT statements. In particular, we show that this structural change provides no clear benefits when using the improved EAT statement. To clearly distinguish the different statements of the EAT, we shall indicate with the subscript $_{\text{DFR16}}$, quantities associated with the original EAT [24] and similarly we shall indicate with the subscript $_{\text{DF18}}$, quantities associated with the newer EAT statement [25].

C.1 Blocked min-tradeoff functions

Let us briefly review the structural modification that was introduced in [26]. Instead of distinguishing the statistics from each interaction separately, rounds are grouped together to form *blocks*. The number of rounds within a block can vary: a new block begins when either a test-round occurs or when the maximum number of rounds permitted within a block, $s_{\max}$, is reached. On expectation there are $\bar{s} = \frac{1-(1-\gamma)^{s_{\max}}}{\gamma}$ rounds within a block. The device-interaction phase of the protocol concludes after some specified number of blocks $m \in \mathbb{N}$ have terminated. We shall use the superscripts ^{R} and ^{B} to indicate whether a

quantity is concerned with the round-by-round or block structured protocols respectively.

The collected information is now defined at the level of blocks and not rounds. In particular, at the end of the i^{th} block the user records some tuple $(\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{X}_i, \boldsymbol{Y}_i, \boldsymbol{C}_i)$, where $(\boldsymbol{A}_i, \boldsymbol{B}_i, \boldsymbol{X}_i, \boldsymbol{Y}_i) \in \mathcal{A}^{s_{\max}} \mathcal{B}^{s_{\max}} \mathcal{X}^{s_{\max}} \mathcal{Y}^{s_{\max}}$ and the scoring alphabet remains the same as in the main text $\boldsymbol{C}_i \in \mathcal{G} \cup \{\perp\}$. The EAT-channels are now defined for each block and the entropy bounding property of min-tradeoff function (cf. (2.3.6) and (2.3.7)) becomes

$$f^B_{\min}(\boldsymbol{p}) \leq \inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i(\sigma)_{\boldsymbol{C}_i} = \tau_{\boldsymbol{p}}} H(\boldsymbol{A}_i \boldsymbol{B}_i | \boldsymbol{X}_i \boldsymbol{Y}_i R')_{\mathcal{N}_i(\sigma)}, \tag{C.1.1}$$

for each $i \in [m]$. The expected frequency distributions for a block's score take the form

$$\boldsymbol{p}^B = \begin{pmatrix} \gamma \bar{s} \boldsymbol{q} \\ (1-\gamma)^{s_{\max}} \end{pmatrix} \tag{C.1.2}$$

for $\boldsymbol{q} \in \mathfrak{Q}_{\mathcal{G}}$.

Lemma C.1 (Blocked variant of Lemma 3.1)**.** *Let $g : \mathfrak{P}_{\mathcal{G}} \to \mathbb{R}$ be an affine function satisfying*

$$g(\boldsymbol{q}) \leq \inf_{\sigma_{R_{i-1}R'} : \mathcal{N}_i^{test}(\sigma)_{\boldsymbol{C}_i} = \tau_{\boldsymbol{q}}} H(\boldsymbol{A}_i \boldsymbol{B}_i | \boldsymbol{X}_i \boldsymbol{Y}_i R')_{\mathcal{N}_i(\sigma)} \tag{C.1.3}$$

for all $\boldsymbol{q} \in \mathfrak{Q}_{\mathcal{G}}$. Then the function $f : \mathfrak{P}_{\mathcal{G} \cup \{\perp\}} \to \mathbb{R}$, defined by

$$\begin{aligned} f(\boldsymbol{\delta}_c) &= \text{Max}[g] + \frac{g(\boldsymbol{\delta}_c) - \text{Max}[g]}{\gamma \bar{s}}, \qquad \forall c \in \mathcal{G}, \\ f(\boldsymbol{\delta}_\perp) &= \text{Max}[g], \end{aligned}$$

is a min-tradeoff function for any EAT-channels implementing Protocol QRE^B. Furthermore, f satisfies the following properties:

$$\begin{aligned} \text{Max}[f] &= \text{Max}[g], \\ \text{Min}_\Sigma[f] &\geq \text{Min}[g], \\ \text{Var}_\Sigma[f] &\leq \frac{(\text{Max}[g] - \text{Min}[g])^2}{\gamma \bar{s}}. \end{aligned}$$

Proof. This follows from replicating the original proof [25] with the block channels decomposed into the testing and generation channels, $\mathcal{N}_i = \gamma \bar{s} \mathcal{N}_i^{\text{test}} + (1 - \gamma \bar{s}) \mathcal{N}_i^{\text{gen}}$. ■

Lemma C.2 (Min-tradeoff construction)**.** *Let $\mathcal{G}$ be a nonlocal game and $k \in \mathbb{N}$. For each $\boldsymbol{\nu} \in \mathfrak{Q}_{\mathcal{G}}^{(k)}$, let $\boldsymbol{\lambda}_{\boldsymbol{\nu}}$ be some feasible point of the dual of Prog.* (2.7.13) *when parametrized by $\boldsymbol{\nu}$ and computed at the k^{th} relaxation level. Furthermore, let $\lambda_{\max} = \max_{c \in \mathcal{G}} \lambda_{\boldsymbol{\nu}}(c)$ and $\lambda_{\min} = \min_{c \in \mathcal{G}} \lambda_{\boldsymbol{\nu}}(c)$. Then, for any set of EAT channels $\{\mathcal{N}_i\}_{i=1}^m$ implementing an instance of Protocol QRE^B with the nonlocal game $\mathcal{G}$, the set of functionals $F^B_{\min}(\mathcal{G}) = \{f_{\boldsymbol{\nu}}(\cdot) \mid \boldsymbol{\nu} \in \mathfrak{Q}_{\mathcal{G}}^{(k)}\}$*

forms a family of min-tradeoff functions, where $f_{\boldsymbol{\nu}} : \mathfrak{P}_{\mathcal{C}} \to \mathbb{R}$ *are defined by*

$$f_{\boldsymbol{\nu}}(\boldsymbol{\delta}_c) := (1-\gamma)\bar{s}\left(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}} \frac{\lambda_{\boldsymbol{\nu}}(c) - (1-\gamma\bar{s})\lambda_{\min}}{\gamma\bar{s}}\right) \quad \text{for } c \in \mathcal{G}, \tag{C.1.4}$$

$$f_{\boldsymbol{\nu}}(\boldsymbol{\delta}_\perp) := (1-\gamma)\bar{s}(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\lambda_{\min}), \tag{C.1.5}$$

where $A_{\boldsymbol{\nu}} = \frac{1}{\ln 2} - \log(\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu})$ *and* $B_{\boldsymbol{\nu}} = \frac{1}{\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{\nu} \ln 2}$.

Moreover, these min-tradeoff functions satisfy the following identities.

- *Maximum:*

$$\mathrm{Max}[f_{\boldsymbol{\nu}}] = (1-\gamma)\bar{s}(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\lambda_{\min}) \tag{C.1.6}$$

- Σ*-Minimum:*

$$\mathrm{Min}_{\Sigma}[f_{\boldsymbol{\nu}}] \geq (1-\gamma)\bar{s}(A_{\boldsymbol{\nu}} - B_{\boldsymbol{\nu}}\lambda_{\max}) \tag{C.1.7}$$

- Σ*-Variance:*

$$\mathrm{Var}_{\Sigma}[f_{\boldsymbol{\nu}}] \leq \frac{(1-\gamma)^2 \bar{s} B_{\boldsymbol{\nu}}^2 (\lambda_{\max} - \lambda_{\min})^2}{\gamma} \tag{C.1.8}$$

Proof. The proof follows the same structure as the proof of Lemma 3.2. The only significant difference is the construction of the function $g : \mathfrak{P}_{\mathcal{G}} \to \mathbb{R}$ satisfying (C.1.3) so we shall explain this part here. Following Appendix B of [26], by repeated application of the chain rule we may decompose a block's entropy as

$$H(\boldsymbol{A}_i\boldsymbol{B}_i|\boldsymbol{X}_i\boldsymbol{Y}_i\boldsymbol{T}_i R')_{\mathcal{N}_i(\sigma)} = \sum_{j=1}^{s_{\max}} (1-\gamma)^{j-1} H(A_{i,j}B_{i,j}|\boldsymbol{X}_i\boldsymbol{Y}_i, \boldsymbol{T}_{i,1}^{\,j-1} = \boldsymbol{0}, \boldsymbol{T}_{i,j}^{\,s_{\max}} \boldsymbol{A}_{i,1}^{\,j-1} \boldsymbol{B}_{i,1}^{\,j-1} R'),$$

where $T_{i,j}$ is the random variable indicating whether a test occurred on the j^{th} round of the i^{th} block. Considering the individual terms within the sum, we can absorb the majority of the side information into some arbitrary quantum register E leaving us with terms of the form

$$(1-\gamma)^{j-1} H(A_{i,j}B_{i,j}|X_{i,j}Y_{i,j}T_{i,j}E).$$

As before, we can use the inequality $H(A|B) \geq H_{\min}(A|B)$ and conditioning on $T_{i,j}$ to lower bound each term in the sum by a feasible point of the semidefinite program,

$$\begin{aligned}
(1-\gamma)^{j-1} H(A_{i,j}B_{i,j}|X_{i,j}Y_{i,j}T_{i,j}E) &= (1-\gamma)^{j-1}\mathbb{P}\left[T_{i,j}=0\right] H(A_{i,j}B_{i,j}|X_{i,j}=\tilde{x}, Y_{i,j}=\tilde{y}, T_{i,j}=0, E) \\
&\quad + (1-\gamma)^{j-1}\mathbb{P}\left[T_{i,j}=1\right] H(A_{i,j}B_{i,j}|X_{i,j}Y_{i,j}T_{i,j}=1\, E) \\
&\geq (1-\gamma)^{j} H(A_{i,j}B_{i,j}|\tilde{x}\,\tilde{y}\,E) \\
&\geq (1-\gamma)^{j} H_{\min}(A_{i,j}B_{i,j}|\tilde{x}\,\tilde{y}\,E) \\
&\geq -(1-\gamma)^{j} \log(\boldsymbol{\lambda}_{\boldsymbol{\nu}} \cdot \boldsymbol{q}_{i,j}),
\end{aligned}$$

where $\boldsymbol{q}_{i,j} \in \mathfrak{Q}_{\mathcal{G}}$ is the expected frequency distribution over the game scores for round j of block i. Noting that $-\log(\cdot)$ of an linear function is convex [82], we can establish a bound on the entire block i through an application of Jensen's inequality[1]

$$(\gamma-1)\sum_{j=1}^{s_{\max}}(1-\gamma)^{j-1}\log(\boldsymbol{\lambda_v}\cdot\boldsymbol{q}_{i,j}) \geq \bar{s}(\gamma-1)\log\left(\boldsymbol{\lambda_v}\cdot\frac{\sum_{j=1}^{s_{\max}}(1-\gamma)^{j-1}\boldsymbol{q}_{i,j}}{\bar{s}}\right)$$

we have used the fact that $\sum_{j\in[s_{\max}]}(1-\gamma)^{j-1}=\bar{s}$. Let $\boldsymbol{q}_i \in \mathfrak{Q}_{\mathcal{G}}$ be the expected nonlocal game score for the i^{th} block conditioned on a test occurring. We may write $\boldsymbol{q}_i$ as

$$\begin{aligned}\boldsymbol{q}_i &= \frac{\sum_{j\in[s_{\max}]}\gamma(1-\gamma)^{j-1}\boldsymbol{q}_{i,j}}{1-(1-\gamma)^{s_{\max}}},\\ &= \frac{\sum_{j\in[s_{\max}]}(1-\gamma)^{j-1}\boldsymbol{q}_{i,j}}{\bar{s}}.\end{aligned}$$

Thus, we have so far established that

$$H(\boldsymbol{A}_i\boldsymbol{B}_i|\boldsymbol{X}_i\boldsymbol{Y}_i\boldsymbol{T}_iR')_{\mathcal{N}_i(\sigma)} \geq \bar{s}(\gamma-1)\log\left(\boldsymbol{\lambda_v}\cdot\boldsymbol{q}_i\right),$$

where $\boldsymbol{q}_i \in \mathfrak{Q}_{\mathcal{G}}$ is the expected frequency distribution over the nonlocal game scores for the i^{th} block. Taking a first order expansion about the point $\boldsymbol{v}$ we arrive at the bound

$$H(\boldsymbol{A}_i\boldsymbol{B}_i|\boldsymbol{X}_i\boldsymbol{Y}_i\boldsymbol{T}_iR')_{\mathcal{N}_i(\sigma)} \geq (1-\gamma)\bar{s}(A_{\boldsymbol{v}}-B_{\boldsymbol{v}}\boldsymbol{\lambda_v}\cdot\boldsymbol{q}_i),$$

where $A_{\boldsymbol{v}}=\frac{1}{\ln 2}-\log(\boldsymbol{\lambda_v}\cdot\boldsymbol{v})$ and $B_{\boldsymbol{v}}=\frac{1}{\boldsymbol{\lambda_v}\cdot\boldsymbol{v}\ln 2}$. Note that the right hand side is a device-independent bound (does not refer to the state and measurements of the system), therefore it is also a bound on $\inf_{\sigma_{R_{i-1}R'}:\mathcal{N}_i^{\text{test}}(\sigma)_{\boldsymbol{C}_i}=\tau_{\boldsymbol{q}}} H(\boldsymbol{A}_i\boldsymbol{B}_i|\boldsymbol{X}_i\boldsymbol{Y}_i\boldsymbol{T}_iR')_{\mathcal{N}_i(\sigma)}$. The prerequisites of Lemma C.1 are now satisfied and the result follows from applying the lemma to the constructed entropy-bounding function above. ■

Remark C.1. The min-tradeoff functions for the blocked protocol are very similar to those of Lemma 3.2. If $\boldsymbol{p} \in \mathfrak{Q}_{\mathcal{G}}$ then

$$f_{\boldsymbol{v}}^{B}(\gamma\bar{s}\boldsymbol{p}, 1-\gamma\bar{s}) = \bar{s}f_{\boldsymbol{v}}^{R}(\gamma\boldsymbol{p}, 1-\gamma).$$

That is, evaluating the corresponding min-tradeoff functions for distributions which respect the structure of the protocols, we get that the blocked function's bound is precisely $\bar{s}$ times the round-by-round bound.

[1] Jensen's inequality [111] states that for a function $\varphi:\mathbb{R}\to\mathbb{R}$, continuous and convex on some interval $I\subseteq\mathbb{R}$, we have

$$\varphi\left(\frac{\sum_{i=1}^{n}\alpha_i x_i}{\sum_{i=1}^{n}\alpha_i}\right) \leq \frac{\sum_{i=1}^{n}\alpha_i\varphi(x_i)}{\sum_{i=1}^{n}\alpha_i},$$

where $x_i \in I$ and $\alpha_i > 0$ for each $i \in [n]$.

C.2 Blocking with the improved second order

The error term in the original EAT bound is

$$\epsilon^R_{\mathrm{DFR16}} := 2(\log(1+2|\mathcal{A}||\mathcal{B}|) + \lceil \|\nabla f_{\min}\|_\infty \rceil)\sqrt{1-2\log(\epsilon_s \epsilon_{\mathrm{EAT}})}. \tag{C.2.1}$$

The disadvantage of using this bound without modification is that the gradient of $f_{\min}$ scales like $1/\gamma$ and so the total error $\sqrt{n}\epsilon^R_{\mathrm{DFR16}}$ scales as $O(\sqrt{n}/\gamma)$. What was noticed in [26] is that by collating the statistics into $m \in \mathbb{N}$ blocks, one can redistribute some of the γ dependence from the gradient term to the $\log(1+2|\mathcal{A}||\mathcal{B}|)$ term such that the total error scales as $O(\sqrt{\frac{n}{\gamma}})$. Moving to the blocked structure and setting $s_{\max} = \lceil 1/\gamma \rceil$, the output alphabets grow exponentially with the size of the blocks and therefore logarithmic term acquires a $1/\gamma$ scaling. In contrast, the scaling of the derivative of the min-tradeoff function is found to be independent of the block size. Fortunately, as our error is now defined for an entire block, we reduce the multiplicative factor on the total error from $\sqrt{n}$ to $\sqrt{m} \approx \sqrt{n/\bar{s}}$. As $\bar{s} \in O(1/\gamma)$, we find that the total error term $\sqrt{m}\epsilon^B_{\mathrm{DFR16}}$ now scales as $\sqrt{n/\gamma}$. By increasing the size of the blocks we have effectively redistributed the testing probability dependence evenly amongst the components of $\epsilon_{\mathrm{DFR16}}$.

In light of this block-induced improvement, it is natural to investigate whether similar advantages can be obtained by applying this technique to the improved EAT statement [25]. Recall the error terms

$$\epsilon^R_V := \frac{\beta \ln 2}{2}\left(\log\left(2|\mathcal{AB}|^2+1\right) + \sqrt{\mathrm{Var}_\Sigma[f]+2}\right)^2, \tag{C.2.2}$$

$$\epsilon^R_K := \frac{\beta^2}{6(1-\beta)^3 \ln 2} 2^{\beta(\log|\mathcal{AB}| + \mathrm{Max}[f] - \mathrm{Min}_\Sigma[f])} \ln^3\left(2^{\log|\mathcal{AB}| + \mathrm{Max}[f] - \mathrm{Min}_\Sigma[f]} + e^2\right) \tag{C.2.3}$$

and

$$\epsilon^R_\Omega := \frac{1}{\beta}(1 - 2\log(p_\Omega \epsilon_s)). \tag{C.2.4}$$

Using the explicit form of the blocked min-tradeoff functions Lemma C.2, we can calculate the approximate size of the error terms for large $s_{\max}$, small γ and $m \approx n^R/\bar{s}$. In particular, we find

$$\begin{aligned} m \cdot \epsilon^B_V &\leq \frac{\beta m \ln 2}{2}\left(\log\left(2|\mathcal{AB}|^{2s_{\max}}+1\right) + \sqrt{\frac{(1-\gamma)^2 \bar{s} B^2_{\boldsymbol{v}} (\lambda_{\max}-\lambda_{\min})^2}{\gamma} + 2}\right)^2 \\ &= O(\beta n s_{\max}) + O(\beta n/\gamma), \end{aligned} \tag{C.2.5}$$

$$\begin{aligned} m \cdot \epsilon^B_K &\leq \frac{m\beta^2}{6(1-\beta)^3 \ln 2} 2^{\beta(\log|\mathcal{AB}|^{s_{\max}} + (1-\gamma)\bar{s} B_{\boldsymbol{v}}(\lambda_{\max}-\lambda_{\min}))} \ln^3\left(2^{\log|\mathcal{AB}|^{s_{\max}} + (1-\gamma)\bar{s} B_{\boldsymbol{v}}(\lambda_{\max}-\lambda_{\min})} + e^2\right) \\ &= \beta^2 2^{O(\beta s_{\max})} O(n s^2_{\max}), \end{aligned} \tag{C.2.6}$$

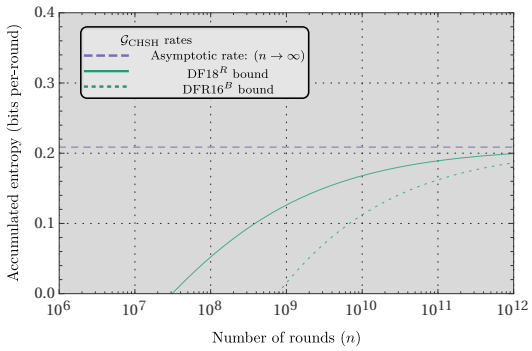
$$\epsilon^B_\Omega = O(1/\beta), \tag{C.2.7}$$

and therefore the total error scales as

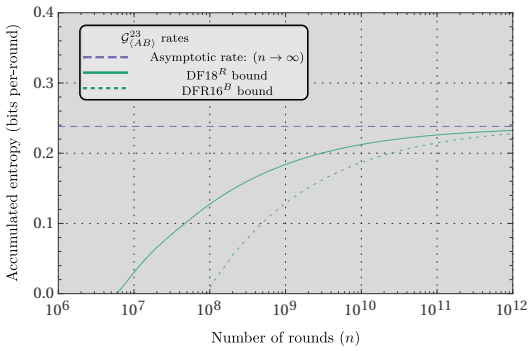
$$O\left(\beta n s_{\max} + \frac{\beta n}{\gamma} + \beta^2 n s_{\max}^2 2^{O(\beta s_{\max})} + \frac{1}{\beta}\right). \tag{C.2.8}$$

In order for ϵ_K^B to have any sensible scaling, we need the exponent to grow no faster than $O(1)$. Combining this with the inverse dependence of β in ϵ_Ω^B, we would like $\beta \approx \frac{\sqrt{\gamma}}{\sqrt{n}\, s_{\max}}$. Such a choice results in the total error scaling as $O\left(s_{\max}\sqrt{n/\gamma}\right)$ which suggests that a large block size is not advantageous with the improved second order statement.

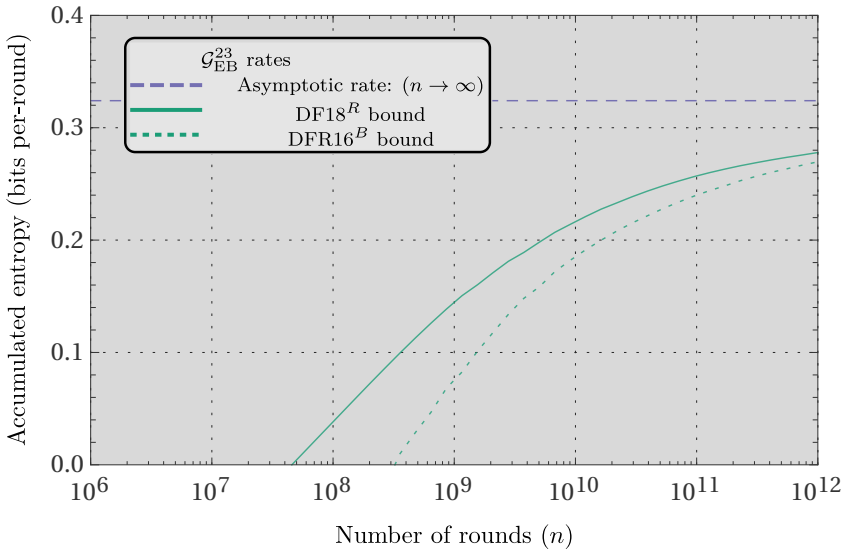
A comparison between the expansion rates obtained when using the improved second order statement [25] and the blocked variant of the original EAT are presented in Fig. C.1. The faster convergence to the asymptotic rate is indicative of the new EAT statement's strength. Additionally, in Fig. C.2 we extend Fig. 4.6 from Sec. 4.3 to compare the four variants of the EAT, i.e. each statement with and without blocking.



(a) Round dependence for $\mathcal{G}_{\text{CHSH}}$



(b) Round dependence for $\mathcal{G}^{23}_{\langle AB \rangle}$



(c) Round dependence for $\mathcal{G}^{23}_{\text{EB}}$

Figure C.1: Comparison of the certifiable accumulation rates using the two different statements of the EAT: DFR16^B [26] and DF18^R (3.2.7). The rates were derived using the following procedure. We assumed a qubit implementation of the protocols with a detection efficiency $\eta = 0.9$, optimizing the state and measurement angles in order to maximise the asymptotic bound. Then, for each value of n an optimization of the min-tradeoff function choice was performed – for the rates calculated using (3.2.7) we also optimized the β parameter at each value of n. To ensure that we approached the asymptotic bound as n increased we set $\gamma = \delta_1 = \cdots = \delta_{|\mathcal{G}|} = n^{-1/3}$ as such a choice provides a constant completeness error across all values of n.

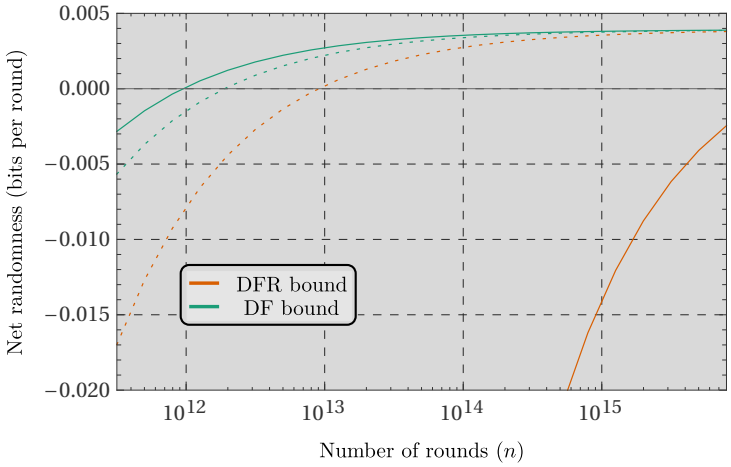


Figure C.2: Comparison between the certifiable accumulation rates of Protocol ARV (see Sec. 4.3) using the blocked and non-blocked variants of the two EAT statements. The dashed lines indicate that the EAT statement was applied to the blocked version of the spot-checking protocol whereas the filled lines indicate the EAT statement was applied round-by-round.

APPENDIX D

ADDITIONAL LEMMAS

Lemma D.1. *Let $X, Y \in \mathfrak{S}_n(\mathbb{R})$ then*

$$\inf_{Y \succeq 0} \mathrm{Tr}[XY] = \begin{cases} 0 & \text{if } X \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \tag{D.0.1}$$

Proof. By the spectral theorem we can rewrite X as $X = \sum_i \lambda_i |x_i\rangle\langle x_i|$.
Case: $X \succeq 0$.
As $X \succeq 0$ we have that $\lambda_i \geq 0$ for each $i \in [n]$ and we may define $X^{1/2} = \sum_i \sqrt{\lambda_i}\, |x_i\rangle\langle x_i|$, which satisfies $X = X^{1/2}X^{1/2}$ and $X^{1/2} \succeq 0$. By the cyclic property of the trace we have $\mathrm{Tr}[XY] = \mathrm{Tr}\left[X^{1/2}YX^{1/2}\right]$ which is non-negative as $X^{1/2}YX^{1/2} \succeq 0$: for $|v\rangle \in \mathbb{R}^n$ we have $\langle v|X^{1/2}YX^{1/2}|v\rangle = \langle w|Y|w\rangle \geq 0$ with $|w\rangle = X^{1/2}|v\rangle$.
Case: $X \not\succeq 0$.
If $X \not\succeq 0$ then at least one of its eigenvalues $\{\lambda_i\}$ is negative. Let λ be one of those negative eigenvalues and define $Y = y\Pi_\lambda$ where Π_λ is the projector onto the eigenspace of λ. For $y \geq 0$ we have $Y \succeq 0$. Moreover, $\mathrm{Tr}[XY] = y\lambda d_\lambda$ where d_λ is the dimension of the eigenspace of λ. Then $\inf_{Y\succeq 0} \mathrm{Tr}[XY] = \inf_{y\geq 0} y\lambda d_\lambda = -\infty$ as $\lambda < 0$. ■

Lemma D.2. *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}^m$, then*

$$\sup_{\boldsymbol{v}\in V} \inf_{\boldsymbol{w}\in W} f(\boldsymbol{v},\boldsymbol{w}) \leq \inf_{\boldsymbol{w}\in W} \sup_{\boldsymbol{v}\in V} f(\boldsymbol{v},\boldsymbol{w}). \tag{D.0.2}$$

Proof. Let $g_-(\boldsymbol{v}) := \inf_{\boldsymbol{w}\in W} f(\boldsymbol{v},\boldsymbol{w})$ and $g_+(\boldsymbol{w}) := \sup_{\boldsymbol{v}\in V} f(\boldsymbol{v},\boldsymbol{w})$. Then, by definition, for any $(\boldsymbol{v},\boldsymbol{w}) \in VW$ we have

$$g_-(\boldsymbol{v}) \leq f(\boldsymbol{v},\boldsymbol{w}) \leq g_+(\boldsymbol{w}).$$

As $g_-(\boldsymbol{v}) \leq g_+(\boldsymbol{w})$ holds for all $(\boldsymbol{v},\boldsymbol{w}) \in VW$, we must also have $g_-(\boldsymbol{v}) \leq \inf_{\boldsymbol{w}\in W} g_+(\boldsymbol{w})$ and in turn $\sup_{\boldsymbol{v}\in V} g_-(\boldsymbol{v}) \leq \inf_{\boldsymbol{w}\in W} g_+(\boldsymbol{w})$. ■

Lemma D.3. *Let* $\boldsymbol{\lambda} \in \mathbb{R}^n$ *and let* $\Lambda = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{\lambda} \cdot \boldsymbol{x} > 0\}$. *Then,*

$$f(\boldsymbol{x}) := -\log(\boldsymbol{\lambda} \cdot \boldsymbol{x}) \tag{D.0.3}$$

is a convex function on Λ.

Proof. Firstly, note that Λ is a convex set. Furthermore, f is a smooth function on Λ and so we can compute its Hessian,

$$H = \frac{1}{\ln 2}\left(\frac{\lambda_i \lambda_j}{(\boldsymbol{\lambda} \cdot \boldsymbol{x})^2}\right)_{i,j}.$$

We may rewrite this as $H = \frac{1}{\ln 2}|v\rangle\langle v|$ with $|v\rangle = \frac{1}{\boldsymbol{\lambda}\cdot\boldsymbol{x}}\sum_i \lambda_i |i\rangle$. Consequently, H is a rank one matrix with a single non-zero eigenvalue $\frac{1}{\ln 2}\| |v\rangle \|^2 = \frac{\sum_i \lambda_i^2}{\ln 2(\boldsymbol{\lambda}\cdot\boldsymbol{x})^2}$. This eigenvalue is manifestly non-negative and thus $H \succeq 0$ and f is a convex function on Λ. ∎

BIBLIOGRAPHY

[1] R. G. Mulgan, "Lot as a democratic device of selection," *The review of politics*, vol. 46, no. 4, pp. 539–560, 1984.

[2] L. Debnath and K. Basu, "A short history of probability theory and its applications," *International Journal of Mathematical Education in Science and Technology*, vol. 46, 01 2015.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[4] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[5] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proceedings 35th annual symposium on foundations of computer science*, pp. 124–134, Ieee, 1994.

[6] M. Wertheimer, "The mathematics community and the nsa," *Notices of the AMS*, vol. 62, no. 2, pp. 165–167, 2015.

[7] D. Mayers and A. Yao, "Quantum cryptography with imperfect apparatus," in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)*, (Los Alamitos, CA, USA), pp. 503–509, IEEE Computer Society, 1998.

[8] A. K. Ekert, "Quantum cryptography based on Bell's theorem," *Physical Review Letters*, vol. 67, no. 6, pp. 661–663, 1991.

[9] J. Barrett, L. Hardy, and A. Kent, "No signalling and quantum key distribution," *Physical Review Letters*, vol. 95, p. 010503, 2005.

[10] S. Pironio, A. Acin, N. Brunner, N. Gisin, S. Massar, and V. Scarani, "Device-independent quantum key distribution secure against collective attacks," *New Journal of Physics*, vol. 11, no. 4, p. 045021, 2009.

[11] R. Colbeck and R. Renner, "Free randomness can be amplified," *Nature Physics*, vol. 8, pp. 450—-454, 2012.

[12] R. Gallego, L. Masanes, G. De La Torre, C. Dhara, L. Aolita, and A. Acin, "Full randomness from arbitrarily deterministic events," *Nature Communications*, vol. 4, p. 3654, 2013.

[13] I. Supić and J. Bowles, "Self-testing of quantum systems: a review," *arXiv preprint arXiv:1904.10042*, 2019.

[14] A. Kent, "Unconditionally secure bit commitment," *Physical Review Letters*, vol. 83, no. 7, pp. 1447–1450, 1999.

[15] J. Silman, A. Chailloux, N. Aharon, I. Kerenidis, S. Pironio, and S. Massar, "Fully distrustful quantum bit commitment and coin flipping," *Phys. Rev. Lett.*, vol. 106, p. 220501, Jun 2011.

[16] R. Colbeck, *Quantum and Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, University of Cambridge, 2007.

[17] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," *Journal of Physics A*, vol. 44, no. 9, p. 095305, 2011.

[18] S. Pironio, A. Acin, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem," *Nature*, vol. 464, pp. 1021–1024, 2010.

[19] S. Pironio and S. Massar, "Security of practical private randomness generation," *Physical Review A*, vol. 87, p. 012336, 2013.

[20] S. Fehr, R. Gelles, and C. Schaffner, "Security and composability of randomness expansion from Bell inequalities," *Physical Review A*, vol. 87, p. 012335, 2013.

[21] C. A. Miller and Y. Shi, "Universal security for randomness expansion from the spot-checking protocol," *arXiv preprint arXiv:1411.6608*, 2014.

[22] C. A. Miller and Y. Shi, "Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, (New York, NY, USA), pp. 417–426, ACM, 2014.

[23] U. Vazirani and T. Vidick, "Certifiable quantum dice or, testable exponential randomness expansion," in *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC-12)*, pp. 61–76, 2012.

[24] F. Dupuis, O. Fawzi, and R. Renner, "Entropy accumulation." arXiv preprint arXiv:1607.01796, 2016.

[25] F. Dupuis and O. Fawzi, "Entropy accumulation with improved second-order term," *IEEE Transactions on Information Theory*, 2019.

[26] R. Arnon-Friedman, R. Renner, and T. Vidick, "Simple and tight device-independent security proofs," *SIAM Journal on Computing*, vol. 48, no. 1, pp. 181–225, 2019.

[27] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, "Practical device-independent quantum cryptography via entropy accumulation," *Nature communications*, vol. 9, no. 1, p. 459, 2018.

[28] E. Knill, Y. Zhang, and H. Fu, "Quantum probability estimation for randomness with quantum side information," *arXiv preprint arXiv:1806.04553*, 2018.

[29] J. Ribeiro, G. Murta, and S. Wehner, "Fully device-independent conference key agreement," *Phys. Rev. A*, vol. 97, p. 022307, Feb 2018.

[30] R. Arnon-Friedman and J.-D. Bancal, "Device-independent certification of one-shot distillable entanglement," *New Journal of Physics*, 2019.

[31] C. Bamps, S. Massar, and S. Pironio, "Device-independent randomness generation with sublinear shared quantum resources," *Quantum*, vol. 2, p. 86, 2018.

[32] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Physical Review Letters*, vol. 23, pp. 880–884, 1969.

[33] N. D. Mermin, "Extreme quantum entanglement in a superposition of macroscopically distinct states," *Physical Review Letters*, vol. 65, no. 15, pp. 1838–1840, 1990.

[34] M. Ardehali, "Bell inequalities with a magnitude of violation that grows exponentially with the number of particles," *Physical Review A*, vol. 46, no. 9, p. 5375, 1992.

[35] A. V. Belinskiĭ and D. N. Klyshko, "Interference of light and bell's theorem," *Physics-Uspekhi*, vol. 36, pp. 653–693, aug 1993.

[36] A. Acín, S. Massar, and S. Pironio, "Randomness versus nonlocality and entanglement," *Phys. Rev. Lett.*, vol. 108, p. 100402, Mar 2012.

[37] O. Nieto-Silleras, S. Pironio, and J. Silman, "Using complete measurement statistics for optimal device-independent randomness evaluation," *New Journal of Physics*, vol. 16, no. 1, p. 013035, 2014.

[38] J.-D. Bancal, L. Sheridan, and V. Scarani, "More randomness from the same data," *New Journal of Physics*, vol. 16, no. 3, p. 033011, 2014.

[39] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4337–4347, 2009.

[40] M. Navascués, S. Pironio, and A. Acín, "Bounding the set of quantum correlations," *Physical Review Letters*, vol. 98, no. 1, p. 010401, 2007.

[41] M. Navascués, S. Pironio, and A. Acín, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations," *New Journal of Physics*, vol. 10, no. 7, p. 073013, 2008.

[42] P. Brown, "Python package for DI protocol development." `https://github.com/peterjbrown519/dirng`, 2018.

[43] P. Brown, S. Ragy, and R. Colbeck, "An adaptive framework for constructing quantum-secure randomness expansion protocols." e-print `https://arxiv.org/abs/1810.13346`, 2018.

[44] W.-Z. Liu, L. Ming-Han, S. Ragy, S.-R. Zhao, B. Bai, Y. Liu, P. Brown, J. Zhang, R. Colbeck, J. Fan, Q. Zhang, and J.-W. Pan, "Experimental demonstration of device-independent randomness expansion." In preparation, 2019.

[45] C. W. Helstrom, *Quantum Detection and Estimation Theory*.
London: Academic Press, 1976.

[46] J. Watrous, *The Theory of Quantum Information*.
Cambridge University Press, 2018.

[47] A. Rényi, "On measures of information and entropy," in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 1, pp. 547–561, 1961.

[48] R. Renner, *Security of Quantum Key Distribution*.
PhD thesis, Swiss Federal Institute of Technology, Zurich, 2005.

[49] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth min- and max-entropies," *IEEE Transactions on information theory*, vol. 56, no. 9, pp. 4674–4681, 2010.

[50] M. Tomamichel, *Quantum Information Processing with Finite Resources: Mathematical Foundations*, vol. 5.
Springer, 2015.

[51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*.
John Wiley and Sons Inc., 2nd ed., 2006.

[52] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*.
Oxford university press, 2013.

[53] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, vol. 23, no. 4, pp. 493–507, 1952.

[54] T. Hagerup and C. Rüb, "A guided tour of chernoff bounds," *Information Processing Letters*, vol. 33, pp. 305–308, 1990.

[55] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.

[56] M. Tomamichel, *A framework for non-asymptotic quantum information theory*.
PhD thesis, Swiss Federal Institute Of Technology, Zürich, 2012.

[57] M. Tomamichel, R. Colbeck, and R. Renner, "A fully quantum asymptotic equipartition property," *IEEE Transactions on information theory*, vol. 55, no. 12, pp. 5840–5847, 2009.

[58] J. von Neumann, "Various techniques used in connection with random digits," *John von Neumann, Collected Works*, vol. 5, pp. 768–770, 1963.

[59] R. Impagliazzo, L. A. Levint, and M. Luby, "Pseudo-random generation from one-way functions," in *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC-89)*, pp. 12–24, 1989.

[60] M. Tomamichel, R. Renner, C. Schaffner, and A. Smith, "Leftover hashing against quantum side information," in *Proceedings of the 2010 IEEE Symposium on Information Theory (ISIT10)*, pp. 2703–2707, 2010.

[61] A. De, C. Portmann, T. Vidick, and R. Renner, "Trevisan's extractor in the presence of quantum side information." e-print `arXiv:0912.5514`, 2009.

[62] R. Konig and R. Renner, "Sampling of min-entropy relative to quantum knowledge," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4760–4787, 2011.

[63] L. Trevisan, "Extractors and pseudorandom generators," *Journal of the ACM*, vol. 48, no. 4, pp. 860–879, 2001.

[64] N. Nisan and A. Ta-Shma, "Extracting randomness: A survey and new constructions," *J. Comput. Syst. Sci.*, vol. 58, no. 1, pp. 148–173, 1999.

[65] A. Fine, "Hidden variables, joint probability, and the bell inequalities," *Physical Review Letters*, vol. 48, no. 5, p. 291, 1982.

[66] D. Avis and K. Fukuda, "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra," *Discrete & Computational Geometry*, vol. 8, no. 3, pp. 295–313, 1992.

[67] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," in *Speakable and unspeakable in quantum mechanics*, ch. 2, Cambridge University Press, 1987.

[68] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?," *Physical Review*, vol. 47, no. 10, pp. 777–780, 1935.

[69] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," *Reviews of Modern Physics*, vol. 86, no. 2, p. 419, 2014.

[70] B. Cirel'son, "Quantum generalizations of Bell's inequality," *Letters in Mathematical Physics*, vol. 4, no. 2, pp. 93–100, 1980.

[71] L. Masanes, A. Acin, and N. Gisin, "General properties of nonsignaling theories," *Physical Review A*, vol. 73, p. 012112, 2006.

[72] A. Acin, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, "Device-independent security of quantum cryptography against collective attacks," *Physical Review Letters*, vol. 98, p. 230501, 2007.

[73] M. Kessler and R. Arnon-Friedman, "Device-independent randomness amplification and privatization," *arXiv preprint arXiv:1705.04148*, 2017.

[74] S. Popescu and D. Rohrlich, "Which states violate Bell's inequality maximally?," *Physics Letters A*, vol. 169, no. 6, pp. 411–414, 1992.

[75] J.-A. Larsson, "Loopholes in bell inequality tests of local realism," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424003, 2014.

[76] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellan, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong loophole-free test of local realism," *Physical Review Letters*, vol. 115, p. 250402, 2015.

[77] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, "Significant-loophole-free test of Bell's theorem with entangled photons," *Physical Review Letters*, vol. 115, p. 250401, 2015.

[78] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abella, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres," *Nature*, vol. 526, pp. 682–686, 2015.

[79] S. Pironio, "Lifting bell inequalities," *Journal of mathematical physics*, vol. 46, no. 6, p. 062112, 2005.

[80] E. Hänggi and R. Renner, "Device-independent quantum key distribution with commuting measurements." e-print `arXiv:1009.1833`, 2010.

[81] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, "Device-independent randomness generation from several Bell estimators," *New Journal of Physics*, vol. 20, no. 2, p. 023049, 2018.

[82] S. Boyd and L. Vandenberghe, *Convex optimization*.
Cambridge university press, 2004.

[83] M. Slater, "Lagrange multipliers revisited," in *Traces and Emergence of Nonlinear Programming*, pp. 293–306, Springer, 2014.

[84] K. Fujisawa, M. Kojima, K. Nakata, and M. Yamashita, "Sdpa (semidefinite programming algorithm) user's manual—version 6.2. 0," *Department of Mathematical and Com-puting Sciences, Tokyo Institute of Technology. Research Reports on Mathematical and Computing Sciences Series B: Operations Research*, 2002.

[85] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[86] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical programming*, vol. 96, no. 2, pp. 293–320, 2003.

[87] S. Pironio, M. Navascués, and A. Acin, "Convergent relaxations of polynomial optimization problems with noncommuting variables," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2157–2180, 2010.

[88] V. Paulsen, *Completely bounded maps and operator algebras*, vol. 78. Cambridge University Press, 2002.

[89] Y. Liu, Q. Zhao, M.-H. Li, J.-Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.-Z. Liu, C. Wu, X. Yuan, *et al.*, "Device-independent quantum random-number generation," *Nature*, vol. 562, no. 7728, p. 548, 2018.

[90] R. Canetti, "Universally composable security: A new paradigm for cryptographic protocols.," in *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science (FOCS-01)*, pp. 136–145, 2001.

[91] B. Pfitzmann and M. Waidner, "A model for asynchronous reactive systems and its application to secure message transmission," in *Proceedings of the 2001 IEEE Symposium on Security and Privacy (SP01)*, (Washington, DC, USA), pp. 184–201, IEEE Computer Society, 2001.

[92] C. Portmann and R. Renner, "Cryptographic security of quantum key distribution," *arXiv preprint arXiv:1409.3525*, 2014.

[93] J. Barrett, R. Colbeck, and A. Kent, "Memory attacks on device-independent quantum cryptography," *Physical Review Letters*, vol. 106, p. 010503, 2013.

[94] T. S. Han and M. Hoshi, "Interval algorithm for random number generation," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 599–611, 1997.

[95] B. Tsirelson, "Some results and problems on quantum Bell-type inequalities," *Hadronic Journal Supplement*, vol. 8, pp. 329–345, 1993.

[96] P. H. Eberhard, "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment," *Physical Review A*, vol. 47, pp. 747–750, 1993.

[97] S. M. Assad, O. Thearle, and P. K. Lam, "Maximizing device-independent randomness from a bell experiment by optimizing the measurement settings," *Physical Review A*, vol. 94, no. 1, p. 012304, 2016.

[98] W. Mauerer, C. Portmann, and V. B. Scholz, "A modular framework for randomness extraction based on Trevisan's construction," *arXiv preprint arXiv:1212.0520*, 2012.

[99] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, *et al.*, "Experimentally generated randomness certified by the impossibility of superluminal signals," *Nature*, vol. 556, no. 7700, p. 223, 2018.

[100] M. Bartock, L. E. Bassham, J. Bienfang, H. Booth, L. T. A. N. Brandao, J. M. Kelsey, P. Kuo, A. Migdall, C. A. Miller, S. W. Nam, and M. Wayne, "Nist randomness beacon."
`https://beacon.nist.gov/home`.

[101] A. Zubkov and A. Serov, "A complete proof of universal inequalities for the distribution function of the binomial law," *Theory of Probability & Its Applications*, vol. 57, no. 3, pp. 539–544, 2013.

[102] Q. Zhang, W.-Z. Liu, and M.-H. Li. private communication, 2019.

[103] F. J. Curchod, M. Johansson, R. Augusiak, M. J. Hoban, P. Wittek, and A. Acín, "Unbounded randomness certification using sequences of measurements," *Physical Review A*, vol. 95, no. 2, p. 020102, 2017.

[104] G. Murta, S. B. van Dam, J. Ribeiro, R. Hanson, and S. Wehner, "Towards a realization of device-independent quantum key distribution," *Quantum Science and Technology*, 2019.

[105] A. Pozas-Kerstjens, R. Rabelo, L. Rudnicki, R. Chaves, D. Cavalcanti, M. Navascues, and A. Acin, "Bounding the sets of classical and quantum correlations in networks," *arXiv preprint arXiv:1904.08943*, 2019.

[106] D. M. Greenberger, M. Horne, and A. Zeilinger, "Going beyond Bell's theorem," in *Bell's Theorem, Quantum Mechanics and Conceptions of the Universe* (M. Kafatos, ed.), pp. 69–72, Dordrecht, The Netherlands: Kluwer Academic, 1989.

[107] A. Winick, N. Lütkenhaus, and P. J. Coles, "Reliable numerical key rates for quantum key distribution," *Quantum*, vol. 2, p. 77, July 2018.

[108] J. Ribeiro, L. P. Thinh, J. Kaniewski, J. Helsen, and S. Wehner, "Device independence for two-party cryptography and position verification with memoryless devices," *Phys. Rev. A*, vol. 97, p. 062307, Jun 2018.

[109] P. Busch, P. Lahti, J.-P. Pellonpää, and K. Ylinen, *Quantum measurement*, vol. 22. Springer, 2016.

[110] P. Wittek, "Algorithm 950: Ncpol2sdpa - sparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables," *ACM Transactions on Mathematical Software (TOMS)*, vol. 41, no. 3, p. 21, 2015.

[111] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Math.*, vol. 30, pp. 175–193, 1906.