Improvements on Device Independent and Semi-Device Independent Protocols of Randomness Expansion

Rutvij Bhavsar

PhD

University of York
Mathematics

August 2023

Abstract

To generate genuine random numbers, random number generators based on quantum theory are essential. However, ensuring that the process used to produce randomness meets desired security standards can pose challenges for traditional quantum random number generators. This thesis delves into Device Independent (DI) and Semi-Device Independent (semi-DI) protocols of randomness expansion, based on a minimal set of experimentally verifiable security assumptions. The security in DI protocols relies on the violation of Bell inequalities, which certify the quantum behavior of devices. The semi-DI protocols discussed in this thesis require the characterization of only one device – a power meter. These protocols exploit the fact that quantum states can be prepared such that they cannot be distinguished with certainty, thereby creating a randomness resource. In this study, we introduce enhanced DI and semi-DI protocols that surpass existing ones in terms of output randomness rate, security, or in some instances, both. Our analysis employs the Entropy Accumulation Theorem (EAT) to determine the extractable randomness for finite rounds. A notable contribution is the introduction of randomness expansion protocols that recycle input randomness, significantly enhancing finite round randomness rates for DI protocols based on the CHSH inequality violation. In the final section of the thesis, we delve into Generalized Probability Theories (GPTs), with a focus on Boxworld, the largest GPT capable of producing correlations consistent with relativity. A tractable criterion for identifying a Boxworld channel is presented.

To Hemi Foi

Contents

Abstract i

Contents v

List of Figures xi

Acknowledgments xiii

Author's declaration xv

1 Introduction to protocols of randomness expansion 1

2 Preliminaries 9

2.1 Optimization problems in quantum information 9

2.2 Semi-Definite Programs 11

2.3 Polynomial optimization problems 13

2.3.1 Proving non-negativity of a polynomial 13

2.3.1.1 Proving if a function is sum of squares 14

2.3.1.2 Using SOS condition to determine positivity 15

2.3.2 Solving an unconstrained polynomial optimization problem 16

2.3.3 Solving a constrained polynomial optimization problem 17

2.4 Operational meaning of the min-entropy 17

2.5 Quantifying randomness in a protocol 21

2.6 Entropy Accumulation Theorem 24

I Device Independent Protocols **27**

3 Introduction to Device Independent Protocols **28**

3.1 Introduction . . . 28
3.2 Bell's theorem and randomness . . . 32
3.3 The progress of DIRNE protocols: a brief review . . . 35
3.4 The significance of various entropic quantities . . . 38

4 Upper bounds on the entropic quantities **41**

4.1 Rates for (generalized) CHSH-based protocols . . . 42
4.2 Simplifying the strategy . . . 44
4.2.1 Reduction to projective measurements . . . 46
4.2.2 Reduction to convex combinations of qubit strategies . . . 47
4.3 Qubit strategies . . . 49
4.4 Reduction to pure states . . . 54
4.5 Simplifications of qubit strategies for specific entropic quantities . . . 55
4.6 H(AB|X=0,Y=0,E) . . . 58
4.7 H(AB|XYE) . . . 59
4.8 H(AB|E) . . . 60
4.9 H(A|X=0,Y=0,E) . . . 60
4.10 H(A|XYE) . . . 63
4.11 H(A|E) . . . 64
4.12 Numerically computing upper bounds on rates . . . 64
4.13 Upper bounds for other entropic quantities . . . 67

5 Lower bounds on the entropies **69**

5.1 Lower bounds . . . 69
5.2 H(A|XYE) . . . 70
5.2.1 Monotonicity properties of the function K . . . 73
5.2.2 Lower bounding the objective function . . . 74
5.2.3 Obtaining a lower bound on $G_{A|XYE}$ . . . 75
5.2.4 Lower bounding the randomness rate . . . 78
5.3 H(AB|X=0, Y=0, E) . . . 79
5.3.1 Reparameterizing the optimisation problem . . . 79
5.3.2 Some simplifications . . . 82

5.3.3 Reduction in parameters . . . 83
5.3.4 Lower bounding the function . . . 86
5.4 H(AB|XYE) . . . 87
5.4.1 Optimization of H(AB|XYE) . . . 88
5.4.2 Partitioning the domain . . . 89
5.5 Monotonicity of rates . . . 91
5.6 Results for the lower bounds . . . 94

6 Results and discussion 96
6.1 CHSH-based spot-checking protocol for randomness expansion . . . 97
6.2 CHSH-based protocols without spot-checking . . . 98
6.3 Discussion . . . 105

II Semi Device Independent Protocols 107

7 Introduction to semi-Device Independent Protocols 109
7.1 Introduction . . . 109
7.2 A brief review of different QRNGs . . . 114
7.3 General semi-Device Independent protocol . . . 116

8 Optimizing the von Neumann entropy 119
8.1 Strategies . . . 119
8.2 Incorporating pre-shared randomness and reduction to projective strategies 121
8.3 Reduction to qubit strategies . . . 125
8.4 Converting the optimization problem to a traditional form . . . 128
8.5 Optimization problem in terms of bounded real variables . . . 129
8.6 Extending the domain . . . 131
8.7 Eliminating one more variable . . . 132
8.8 Computing the optimization problem over grids . . . 133
8.9 Converting the optimization problem to a polynomial optimization problem 134
8.10 Taking the convex lower-bound of the rate . . . 136

9 Results and discussion 138
9.1 Recycled inputs protocol . . . 139
9.2 A protocol to generate private randomness . . . 140

9.3 Rates of the protocol . . . 142
9.4 Results . . . 143
9.5 Discussion . . . 146

10 Conclusion and outlook **147**

III Generalized Probability Theories **151**

11 Generalized Probability Theories **153**
11.1 Introduction . . . 153
11.2 The GPT framework: A concise literature overview . . . 155
11.3 State spaces . . . 157
11.4 Transformations . . . 159
11.5 Effects . . . 161
11.6 GPTs as diagrams . . . 163

12 Channels in Boxworld **165**
12.1 Notations and definitions . . . 165
12.2 Redefining Boxworld state space . . . 169
12.3 Some examples and properties of completely positive map . . . 170
12.4 Positive maps and completely positive maps in Boxworld . . . 173
12.5 Conclusion . . . 176

13 Discussion and conclusion **178**

A Appendix for Device Independent Protocols **181**
A.1 Upper bounding the derivatives - H(A|XYE) . . . 181
A.2 Upper bounding the derivatives - H(AB|00E) . . . 183
A.3 Appendix for H(AB|XYE) . . . 185
A.3.1 General result on Grids . . . 185
A.4 Usage of EAT for different protocols . . . 187
A.4.1 Protocol with recycled input randomness (Protocol 3) . . . 187
A.4.2 Spot-checking CHSH protocol (Protocol 1) . . . 188
A.4.3 Protocol with biased local random numbers (Protocol 2) . . . 189
A.4.3.1 Deriving the min-tradeoff function . . . 189
A.4.3.2 Completeness error . . . 191

A.4.4 Error parameters 192
A.4.5 Application to $H(AB|E)$ and $H(A|E)$ 193
A.5 Discussion of composability 193

B Appendix for semi-Device Independent protocols **195**
B.1 Useful claims 195
B.2 Elminating some parameters 197
B.3 Monotonicity of rates 198

References **200**

List of Figures

1.1 A typical protocol of randomness expansion. 5

2.1 A schematic depiction of the sets SOS and POS. 16
2.2 A diagram depicting the scenario in which an adversary is trying to guess the output random variable of a protocol. 19
2.3 A protocol of randomness expansion in terms of sub-protocols. 22

3.1 A diagram depicting a typical Bell scenario involving two parties. 33
3.2 A diagram depicting an informal summary of Bell's theorem. 34

4.1 Graphs of the rates for (a) the one-sided and (b) the two-sided randomness with uniformly chosen inputs. Each of these curves has a non-linear part and the blue curves do not have a linear part. 66
4.2 (a) Two-sided and (b) one-sided entropy curves conditioned on X, Y and E with uniform input distribution. 67
4.3 (a) Two-sided and (b) one-sided entropy curves conditioned on the system E with uniform input distribution. 68

5.1 Graphs of the conjectured rates and lower bounds for various DIRNE protocols assuming that Alice and Bob share qubits. 95
5.2 Graphs for lower bounds on randomness rates for protocols that recycles input randomness. 95

6.1 Graphs of the net rate of certifiable randomness according to the EAT for (a) the spot checking protocol, (b) the protocol with recycled input randomness, and (c) the protocol with biased local random number generators, showing the variation with the number of rounds for three different scores. 103

6.2 Graphs of the net rate of certifiable randomness according to the EAT for (a) the spot checking protocol, (b) the protocol with recycled input randomness, and (c) the protocol with biased local random number generators, showing the variation with the CHSH score . The round numbers, are indicated in the legend. 104

7.1 A schematic diagram of our semi-DI protocol. 116

9.1 Lower bound of the function G. 144
9.2 Asymptotic rates in the generate round of our semi-DI protocol. 144

Acknowledgments

My deepest thanks go to my supervisor, Prof. Roger Colbeck, whose mentorship during my PhD journey has been invaluable. His profound understanding of mathematics and physics has significantly deepened my own knowledge and academic growth. I am especially grateful for his role in making my PhD journey rewarding and truly delightful, even against the backdrop of the challenges presented by the COVID-19 pandemic.

Thanks are due to Dr. Stefan Weigert and Dr. Gustav Delius for their consistent mentorship and guidance as members of my TAP. I also thank Prof. Serge Massar and Dr. Stefan Weigert for taking time to review my thesis and provide a detailed feedback on the thesis. I must acknowledge the enriching discussions with my collaborators Dr. Hamid Tebyanian and Dr. Sammy Raggy.

I am extremely grateful for the insights and camaraderie of fellow PhD students in Quantum Information theory group including Vilasini, Vincenzo, Lewis, Shashaank, Max, Vasilis, Alistair, and Kuntal. Special thanks to Jintao, Lewis, Vincenzo, Ambroise, and Hamid for reviewing sections of this thesis and helping with LaTeX issues.

Immense gratitude is extended to my family-my parents, Manisha and Vihang, along with my grandparents, Sarojben, Chandrikaben, Natvarlal, and Jagdishchandra-who have been pillars of support and wisdom throughout this journey. I owe special thanks to Jintao for his unwavering support and patience. I am also indebted to my circle of friends, including Alba, Diego, Eva, Brad, Peiyun, David, Matt, Macey, Cordelia, Jenny, Beth, Ding, Ambroise, Andrew, Esther, Samantha, Sam, Jack, Jade, Guy, Ross, Emily, Simen, Firat, Vidul, Sapna, Drishti, Hardik, Akshita, Nekhel, Anjali, Ayan, Shefali, Ashvani, Shubham, Laura, Shreya, Aditya, Khushboo, and Varun, whose camaraderie and encouragement have significantly enriched my PhD experience.

Finally, I acknowledge the Quantum Communications Hub and the WW Smith Fund for financial support, which was instrumental in completing my PhD.

Author's declaration

I declare that the work presented in this thesis, except where otherwise stated, is based on my own research carried out at the University of York under the supervision of Prof Roger Colbeck and has not been submitted previously for any degree at this or any other university. Sources are acknowledged by explicit references. The sections on randomness expansion protocols are largely based on the research works stated below:

Chapters 3, 4, 5 and 6 are based on the following joint works:
(1) R. Bhavsar, S. Ragy, and R. Colbeck. "Improved device-independent randomness expansion rates using two sided randomness". New Journal of Physics 25.9 (2023): 093035.
(2) R. Bhavsar and R. Colbeck. "Jordan's Lemma assisted lower-bounds on von-Neumann entropies for DIRNE/DIQKD protocols". In preparation.
The works in Chapters 7, 8 and 9 are based on the joint work:
(3) R. Bhavsar, H. Tebyanian, and R. Colbeck. "A Semi-Device independent protocol for randomness expansion". In preparation.

Parts of Chapter 4 including 4.2, 4.3, 4.4, 4.5 and 4.9 are extensions of the work:
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 11.4 (2009): 045021.
A discussion of the relationship of this work and the above work can be found in Chapter 4.

I have contributed to **all** aspects of research works (1), (2), and (3). Portions of this work were proofread with the assistance of online tools, Grammarly and ChatGPT, as

well as the valuable input from Mr Jintao Shuai, Mr Lewis Wooltorton, and Mr Vincenzo Fiorentino. All such assistance is limited to proofreading and identification of grammatical errors. I declare that no violations of the University's Guidance on Proofreading and Editing have been made.

— 1 —

Introduction to protocols of randomness expansion

Random numbers have a wide range of applications including cryptography, scientific experiments, games, lotteries, and gambling. Their use in these contexts is to ensure fairness and unpredictability of outcomes.

The process of generating random numbers is often associated with a certain level of suspicion. This suspicion stems from the question of whether the numbers generated are truly random. If the randomness is compromised, the fairness of the numbers, which is one of their strengths, could be significantly weakened. For example, consider a bet based on the outcome of a coin toss. If one participant, a resourceful scientist, could account for all the forces acting on the coin, they could predict the outcome of the toss. This would make the bet unfair, even though it appears fair on the surface. This scenario, although impractical to achieve using everyday technology, demonstrates that a process that appears random to one party may be completely deterministic to another party. Therefore, any process that appears to be random may not be truly random, and could be used to gain an unfair advantage.

With the advent of advanced techniques such as machine learning, which can analyze large sets of data and predict outcomes, the vulnerability of randomness-based systems increases. Everyday activities such as making payments with a card or conducting transactions online depend on the security provided by random numbers. The potential gain from cracking such systems could be substantial, thus providing significant incentives for individuals or organizations to attempt guessing these random numbers. Similarly, in our data-driven world, where access to data is often considered key, governments and companies may be incentivized to break such random-number based security protocols

to gain unauthorized access to private information.

A random number generator (RNG) is a device engineered to produce a sequence of numbers following a uniform probability distribution. However, as discussed above, this definition does not entirely capture the complexities and practical necessities associated with RNGs. A desired characteristic of RNGs is their ability to generate"genuine" or"true" randomness. This means that the sequence of random numbers they produce is not only unpredictable to the party using the RNG but also inaccessible or unknown to any third party, including the manufacturer of the device. In other words, an RNG should be capable of producing a sequence of numbers following a uniform probability distribution for every possible agent who wants to determine it. Such a robust definition is critical in various applications that utilize random numbers, especially when used in applications such as in cryptographic protocols.

Based on our current understanding of physics, it is not possible to attain randomness via classical processes. The reason being, classical systems are fundamentally deterministic, meaning an extremely powerful adversary could, in principle, determine the outcomes of a classical process with certainty. Moreover, even incomplete knowledge regarding the mechanism of randomness generation could be used to make informed guesses about the generated numbers. Consequently, we should focus on quantum processes as potential candidates for constructing a reliable RNG.

Realizing such a robust RNG presents a significant challenge even when using a quantum process. Consider, for example, a RNG designed to use a single photon passing through a 50:50 beam splitter. Theoretically, according to quantum mechanics, performing this process would result in a photon either going to one port of the beam splitter or the other with a 50 percent probability. Therefore, in principle, this process can be used as a quantum random number generator. However, in the real world, there are no guarantees that the outputs are genuinely uniformly distributed, due to inevitable practical imperfections in the beam splitter and the laser source.

Moreover, if complex processes are incorporated within the RNG, the entire device must be meticulously modeled for two primary reasons. First, this is to ensure there has been no interference or tampering by an adversary, which might even be the manufacturer of the RNG. Second, detailed modelling of each individual component of the RNG is crucial to identify and account for any device imperfections. Any imperfections in the RNG could skew the distribution of the generated numbers or introduce predictability, both of which would undermine the randomness and security of the RNG.

Device Independent (DI) protocols aim to get around this. They base security on minimal

and easily verifiable details about pairs of devices and without making any assumption on the inner working of the devices. Instead, they rely on verifying that the input-output statistics of the devices exhibit non-locality [1–3]. In essence, Bell's theorem is used to assure the privacy of the outputs.

Bell's theorem implies that if two (or more) devices that cannot communicate are supplied with random inputs, and violate a Bell inequality they must be generating randomness *no matter what their internal operations are*. This suggests using a Bell inequality violation to construct a Device Independent randomness expansion protocol. In such a protocol, inputs are repeatedly provided to two separated (non-communicating) devices and their outputs are stored. We call the inputs X_i and Y_i and the respective outputs A_i and B_i, where i runs from 1 to n (see chapter 6 for examples of such protocols). After making n inputs we can estimate the average value of some Bell inequality; if there is no violation (or the violation is too small) then the protocol aborts. If the violation is large enough, then the raw outputs are run through an extractor to generate the final output randomness.

Unfortunately, the practical application of DIRNE protocols poses challenges, primarily because achieving violations of Bell inequalities in experiments is very difficult. Although Bell's theorem was introduced in the 1960s, it took decades before the first loophole-free Bell inequality violations were experimentally verified [4–6], and even these showed only modest deviations. For example, a 2021 study [7] reported a CHSH score of 0.752484, a value merely 0.024 above the local bound of 0.75. This violation is significantly smaller than the maximum possible violation, which is roughly 0.853. As we will delve into later in this thesis, a high CHSH score is crucial to implement a robust (CHSH-based) DIRNE protocol.

Although achieving a Bell violation experimentally currently poses a challenge for DI protocols, significant advancements have been made in this area. For example, the violation of the local bound increased from 0.00027 in 2018 as reported by [8], to 0.024 in less than three years. As technology evolves to accommodate these needs, it is worthwhile to explore scenarios where we can trust parts of the system. Such protocols are termed semi-Device Independent (semi-DI) protocols, which are simpler to implement experimentally compared to fully device independent protocols.

A novel randomness expansion protocol was recently proposed, which operates based on energy and overlap bounds [9]. Fundamentally, this protocol is based on a prepare-and-measure scenario consisting of a source and a measurement device. In the protocol, a source generates specific states, either ρ_0 or ρ_1, and the measurement device is used to

determine if ρ_0 or ρ_1 was sent. The states ρ_0 and ρ_1 are prepared in such a way that they are near-identical (yet not entirely identical) to the unique ground state of the system (the vacuum state in the case of a laser being the source). Being almost identical, these states cannot be perfectly distinguished according to quantum mechanics[1], which forms the basis for randomness generation in this protocol.

However, a major assumption in this scenario is that the prepared states are near identical to the ground state. Unfortunately, this assumption cannot be verified in a device independent manner. To address this, the protocol assumes the availability of a trusted power meter, which can measure the energy of the states prepared by the source. For the protocol that we describe in this work, the power meter is the only trusted component in the protocol. Thus, this approach is appealing because it only requires the characterization of a single component of the protocol.

Similar to the DI protocols, this protocol should abort if the outputs do not contain sufficient extractable randomness. In particular, the protocol should abort if either of two conditions is met: the energy of the states is too high, or the detector is unable to distinguish the states ρ_0 and ρ_1 with the desired accuracy.

In this work, we study randomness expansion protocols based on the key ideas outlined above. These protocols make minimal assumptions on the inner workings of the devices used to generate randomness.

While asserting that the protocols discussed in this thesis are secure, we assume that the quantum theory is correct and complete [10]. However, it is worth noting that the security of the Device Independent protocols does not rely solely on this assumption; their security has also been established under a significantly weaker condition: that the eavesdropper is bound only by the non-signalling principle [11] (see Chapter 11 for an explanation. Here, the non-signalling principle simply means that no superluminal signalling is possible, without necessarily asserting the correctness of the quantum theory). We also assume that the protocol is being carried out in a secure laboratory from which no information can leak out. It is crucial that humans conducting in the protocol do not cause any data leaks. In other words, having a laboratory that is well-shielded from the outside world is crucial for the security of the protocol. These assumptions are however essential for any protocol of randomness expansion. If such a protective measure is not in place, any randomness expansion protocol can be compromised, regardless of the strategy used.

[1] Note that when an arbitrarily large number of copies of two quantum states are available, perfect distinction between them becomes possible. However, here we consider the case when only a single copy of either of the two states is sent to the measurement device at a given time.

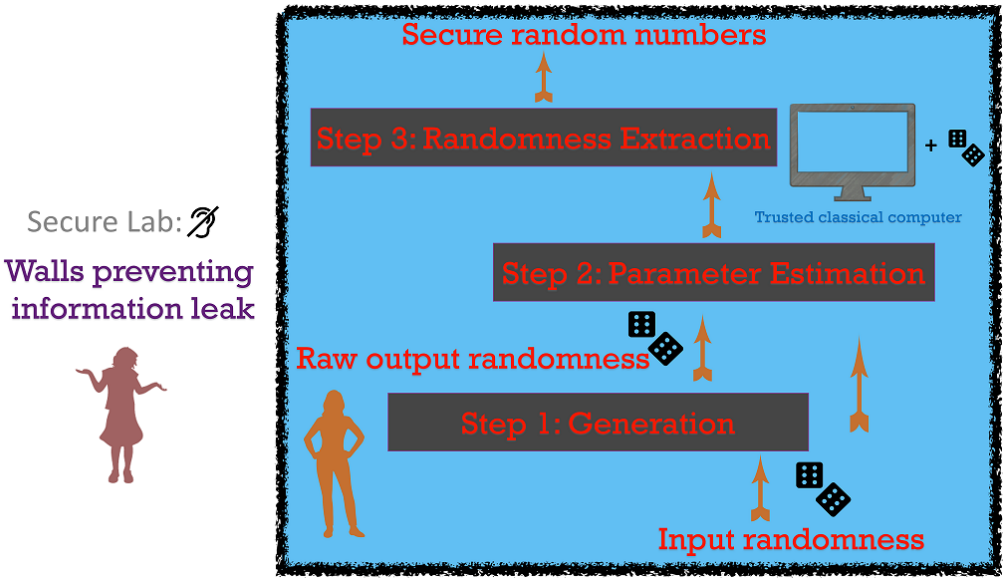


Figure 1.1: A typical protocol consists of three steps - randomness generation, parameter estimation and randomness extraction. Randomness extraction is typically done using a secure classical computer, using known algorithms. The lab is assumed to be secure from which no information can leak out.

Typically, Device Independent and semi-Device Independent protocols are carried out in three different stages as shown in Figure 1.1. These different stages are: generation, parameter estimation, and randomness extraction. The generation stage is the main part of the protocol where a certain sub-protocol is repeatedly executed. In the context of DI protocols, this step involves performing a CHSH test, while in the semi-DI protocol, it is a "prepare and measure round".

The parameter estimation stage involves analyzing the input and output statistics collected in the generation stage. The protocol is aborted if these statistics are not suitable. In DI protocols, the parameter estimation step involves computation of the Bell score. If the observed Bell score is too low (for example a CHSH score less than $3/4$), then the protocol is aborted. If the protocol does not abort, then the amount of uniform randomness that can be extracted from the outputs of the protocol is computed using *only* the input-output statistics of the protocol (such as the Bell score).

The final stage is the randomness extraction stage, where the random strings obtained in the generation stage are processed to produce a string of uniformly distributed random numbers that are secure against any adversary. This stage also consumes some randomness. There is a rich literature on randomness extraction (see for example [12]);

however, this part of the protocol is beyond the scope of this thesis.
After developing these protocols, the primary challenge lies in calculating the amount of extractable randomness as a function of the observed statistics. The difficulty of this task is amplified by the lack of structure in the problem: minimal assumptions are made on how the devices operate, requiring accounting for arbitrary pre-shared entanglement, arbitrary measurements, and potentially adaptive strategies between the rounds. Two techniques to address this exist in the literature: the quantum probability estimation framework [13] and the entropy accumulation theorem (EAT) [14, 15]. In this work, we employ the latter. The EAT, informally speaking, suggests that the amount of extractable randomness in the 'n' bits long full string of outputs is predominantly 'n' times the von Neumann entropy of a single-round strategy that would yield the observed score if used in an independent and identically distributed (i.i.d.) way. This implies that if we can solve the problem for an i.i.d. adversary, we can obtain a bound for the general case.
Consequently, the task of determining the randomness rate (i.e. amount of randomness generated per round) in a DI and our semi-DI protocol reduces to calculating the least value of the von Neumann entropy in a single representative round of the protocol. Computing lower bounds on the von Neumann entropy generally necessitates sophisticated mathematical techniques in optimization theory, especially in the areas of convex optimization. In this work, we develop techniques that can be employed to compute lower bounds on the conditional von Neumann entropy when the inputs and outputs in each round of the protocol are binary. Essentially, this is a situation where we can apply Jordan's Lemma, leading to a significant reduction in the complexity of the problem.
This thesis focuses on the first two stages of the DI and semi-DI protocol discussed above: the generation stage and the parameter estimation stage. Our work focuses on two main aspects:

- Providing new protocols for randomness expansion that use the same setup as traditional protocols, but promise more efficient and secure ways of generating randomness (for example: by reducing experimental assumptions).
- Finding mathematical techniques to compute bounds on randomness rates in DI and semi-DI protocols is introduced.

We now present a brief outline of the thesis:
The subsequent chapter (Chapter 2) discusses some optimization techniques, including the entropy accumulation theorem, which are crucial in randomness expansion protocols. We then move to Chapter 3, where we provide a detailed introduction to DI protocols for

randomness expansion. We also deliberate on various entropic quantities of interest in the context of DI protocols for Randomness Expansion and Quantum Key Distribution. Next, in Chapter 4, we utilize Jordan's Lemma to calculate numerical upper bounds on these entropies. In the following chapter (Chapter 5), we derive the lower bounds for these entropies.

The final chapter in the section on DI protocols for randomness expansion (Chapter 6) presents all the protocols and exhibits the randomness rates achieved for finite rounds using the entropy accumulation theorem.

Following our discussion on DI protocols, we shift our focus to semi-DI protocols. We divide our work into 3 parts. The first part introduces basic framework for the semi-DI (Chapter 7) and explores different protocols. Subsequently, in the next chapter (Chapter 8), we calculate the rates for these semi-DI protocols and in the final chapter on semi-DI protocols, we discuss two different protocols along with results and discussion. We summarize our results on randomness expansion protocols.

Upon concluding our discussion on randomness expansion protocols, we transition to the part of the thesis dealing with Generalized Probability Theories (GPTs) (Chapters 11 and 12).

— 2 —

Preliminaries

2.1 OPTIMIZATION PROBLEMS IN QUANTUM INFORMATION

The theory of optimization problems is a vast field that has been extensively studied in various disciplines such as mathematics, computer science, physics, and economics. As we shall explore further in this chapter, these optimization problems play a key role in quantum information theory as well. The main goal for the study of optimization theory is to compute the value for the following:

$$\min_{x \in \mathcal{D}} g(x). \tag{2.1}$$

In this case, $\mathcal{D}$ is the domain of the optimization problem often also referred to as the feasible set, while $g : \mathcal{D} \mapsto \mathbb{R}$ is known as the objective function. Because of the abstract nature of the problem, there are no universal algorithms available to solve a given optimization problem, as g could be any function and $\mathcal{D}$ could be any set.
Nevertheless, there exists a particular category of optimization problems, known as convex optimization problems, that have gained considerable interest due to their simplicity and the fact that they often have a unique solution. The optimization problem 2.1 is a convex optimization problem if the objective function g is a convex function and the domain $\mathcal{D}$ is a convex set.
Let's define some key terms related to convex optimization:

Definition 1 (Convex Set)**.** A set $\mathcal{D}$ is a convex set if for any points x and y in $\mathcal{D}$, the point $\mu x + (1 - \mu)y$ also belongs to $\mathcal{D}$ for every $\mu \in [0, 1]$.

Definition 2 (Convex Function)**.** Let $\mathcal{D}$ be a convex set, and $g : \mathcal{D} \mapsto \mathbb{R}$ be a function. The function g is convex if, for any points x and y in $\mathcal{D}$ and any value $\mu \in [0, 1]$,

$g(\mu x + (1-\mu)y) \leq \mu g(x) + (1-\mu)g(y)$.

Definition 3 (Concave function)**.** A function f is called concave if $-f$ is convex.

Convex optimization problems frequently appear in quantum information theory. This is primarily due to two fundamental convex sets in quantum theory: the set of states (state space) and the set of effects (effect space).
In the quantum theory, a physical system is associated with a Hilbert space $\mathcal{H}$. A state is a non-negative linear map $\rho : \mathcal{H} \mapsto \mathcal{H}$, with a bounded trace $\mathrm{tr}(\rho) \leq 1$. We denote the set of all states on $\mathcal{H}$ by the set $\mathcal{S}(\mathcal{H})$. This set can be easily shown to be a convex set. Meanwhile, an effect in quantum theory is defined as a linear map $E : \mathcal{S}(\mathcal{H}) \mapsto \mathbb{R}$ such that, for any state $\rho \in \mathcal{S}(\mathcal{H})$, $E(\rho) = \mathrm{tr}(E\rho) \in [0, 1]$. The set of all effects is also easily shown to be a convex set. We will further revisit and generalize these concepts in the chapter discussing Generalized Probability Theories (GPTs) later in Chapter 11.
For completeness, we define a POVM (Positive Operator-Valued Measure) as a collection of effects $\{E_i\}_{i=1}^{N}$ that sum up to the identity $\mathbb{1}$. A POVM describes a measurement in quantum theory, where the probability of obtaining outcome i in an experiment is given by $P[i|\rho] = \mathrm{tr}(E_i\rho)$.
Various convex and concave functions arise naturally in quantum theory. One example of a convex function on the state space is the probability of an outcome corresponding to an effect, denoted as $E(\cdot) := \mathrm{tr}(E\cdot)$. Another concave function that features in quantum information theory is the von Neumann entropy, denoted by $H(\rho)$. Hence, the theory of convex optimization is frequently used in the study of quantum information theory. In the forthcoming sections, we will encounter a range of intriguing convex optimization problems within the context of quantum information theory. The scope of the thesis for discussion of the role of optimization problems in quantum information is limited, and those unfamiliar with the topic are encouraged to go through excellent lecture notes for a good overview of this topic [16].
As will become clear at the end of this chapter, a key quantity of interest that will appear as the objective function in most of the optimization problems in this thesis is the conditional von Neumann entropy:

Definition 4 (von Neumann entropy)**.** The von Neumann entropy of a state $\rho \in \mathcal{S}(\mathcal{H}_A)$ is defined as

$$H(\rho) \equiv H(A)_\rho := -\operatorname{tr}\left(\rho \log_2 \rho\right). \tag{2.2}$$

Here $\log_2$ is the matrix logarithm with base 2.

Definition 5 (conditional von Neumann entropy)**.** For a state $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ the conditional von Neumann entropy $H(A|B)_{\rho_{AB}}$ is given by

$$H(A|B)_{\rho_{AB}} := H(AB)_{\rho_{AB}} - H(B)_{\rho_B}, \tag{2.3}$$

where $\rho_B := \mathrm{tr}_A(\rho_{AB}) \in \mathcal{S}(\mathcal{H}_B)$ is the state on the subsystem B.

2.2 SEMI-DEFINITE PROGRAMS

Even though a general convex optimization problem often holds the appeal of having a unique solution, this solution can be quite challenging to find. The complexity stems from the lack of efficient numerical algorithms capable of reliably solving such general convex optimization problems. Nevertheless, certain optimization problems that we frequently encounter belong to specific classes that can be solved using efficient numerical algorithms. Prominent examples of such problems include Linear Programs and Semi-Definite Programs (SDPs).

Definition 6 (Semi-Definite Program)**.** An optimization problem is an SDP if it is of the following form

$$\begin{aligned} g = \quad & \inf_{X \in \mathbb{H}^m} \mathrm{tr}(CX) \\ \text{s.t.} \quad & X \succeq 0 \\ & \forall i: \quad \mathcal{A}_i(X) = B_i. \end{aligned} \tag{2.4}$$

where $C, X \in \mathbb{H}^m$ (the set of $m \times m$ Hermitian matrices) , $B_i \in \mathbb{H}^{n_i}$ and $\mathcal{A}_i : \mathbb{H}^m \mapsto \mathbb{H}^{n_i}$ is a linear map.

As stated above, efficient numerical algorithms exist that can solve SDPs numerically. Popular software packages like Mosek [17] and CVXPY [18] have integrated these algorithms in a very user-friendly fashion. These algorithms and details on how to program them efficiently are beyond the scope of the thesis.
SDPs are particularly relevant to the study of quantum information theory, as we often need to maximize or minimize a linear function involving arbitrary quantum states or effects. Consider, for example, the problem of operationally distinguishing two quantum states. Suppose we are given a uniform mixture of two states, ρ_1 and ρ_2, and we aim to perform a single measurement to determine whether ρ_1 or ρ_2 was prepared. We are

interested in determining the best possible strategy that allows us to distinguish these states. This problem can be directly formulated as an optimization problem.
To do so, we establish a decision rule by employing a two-outcome measurement $\{E_1, E_2 = \mathbb{1} - E_1\}$. If we observe the outcome corresponding to E_1, we guess that ρ_1 was prepared, and vice versa. The success probability p_{succ} of our strategy can be computed as:

$$\begin{aligned} p_{\text{succ}} &= \frac{1}{2}\mathbb{P}[E_1|\rho_1] + \frac{1}{2}\mathbb{P}[E_2|\rho_2] \\ &= \frac{1}{2}\left(\text{tr}\left(E_1\rho_1\right) + \text{tr}\left(E_2\rho_2\right)\right) \\ &= \frac{1}{2} + \frac{1}{2}\,\text{tr}\left(E_1(\rho_1 - \rho_2)\right). \end{aligned} \tag{2.5}$$

Here $\mathbb{P}[E_i|\rho_i]$ is the probability of observing the outcome $i \in \{0, 1\}$ given that the state ρ_i has been prepared. In order to find the optimal measurement, we would like to maximize the success probability. As $\{E_1, E_2\}$ is a POVM, we have the constraint $0 \leq E_1 \leq \mathbb{1}$. This means that our best chance of the guessing the correct state in this scenario is given by the optimization problem

$$\begin{aligned} p_{\text{succ}}^{\max} = \quad & \sup\left(\frac{1}{2} + \frac{1}{2}\,\text{tr}\left(E_1(\rho_1 - \rho_2)\right)\right) \\ \text{s.t.} \quad & E_1 \succeq 0 \\ & \mathbb{1} - E_1 \succeq 0. \end{aligned} \tag{2.6}$$

The above problem can be easily identified as a SDP. Notably, this optimization problem also has an analytical solution known as the Holevo-Helstrom theorem [19, 20], which states that the maximum success probability is given by:

$$p_{\text{succ}}^{\max} = \frac{1}{2} + \frac{1}{4}||\rho_1 - \rho_2||_1, \tag{2.7}$$

where $||.||_1$ is the trace norm defined as

$$||\tau||_1 := \text{tr}\left(\sqrt{\tau^2}\right).$$

The POVM $\{E_1, E_2\}$ which leads to the maximum success probability is the POVM $\{P_+, P_-\}$ where P_+ and P_- are the positive part and the negative part of the operator $\rho - \sigma$[1]. In the next sections, we will explore the different SDPs that can arise in studying protocols of randomness expansion.

[1]Every Hermitian operator X has a unique deposition $X = P_+ - P_-$, where the operators $P_+ \succeq 0$ and $P_- \succeq 0$ are the called the positive and negative part of X respectively. If the operator X is traceless then $P_+ + P_- = \mathbb{1}$.

2.3 POLYNOMIAL OPTIMIZATION PROBLEMS

Another interesting set of optimization problems are the polynomial optimization problems. The polynomial optimization problems take the form

$$\begin{aligned} \min \quad & p(x_1, \cdots x_n) \\ \text{s.t.} \quad & g_i(x_1, \cdots x_n) \geq 0 \\ & h_j(x_1, \cdots x_n) = 0, \end{aligned} \tag{2.8}$$

with p, q_i and h_j all being polynomials of degree less than or equal to $d \in \mathbb{N}$ and $x_1, \cdots x_n$ being real valued variables - i.e.

$$\begin{aligned} p &\in \mathbb{R}[x_1, \cdots, x_n]_{\leq d}, \\ g_i &\in \mathbb{R}[x_1, \cdots, x_n]_{\leq d}, \\ h_j &\in \mathbb{R}[x_1, \cdots, x_n]_{\leq d}. \end{aligned}$$

Note that such optimization problems are not necessarily convex. Due to the lack of structure in such problems, there are no direct techniques for efficiently solving them. However, we shall demonstrate that any polynomial optimization problem can be cast into a converging sequence of SDPs, each of which can be solved using established algorithms. This approach allows us to compute reliable upper and lower bounds for these optimization problems. It is important to mention, though, as we go higher up in the hierarchy, the SDPs become increasingly more time and resource consuming to solve. Our discussion in this section is primarily based on the lectures on semi-definite programming by Prof. Hamza Fawzi (refer to [21] for the slides). We have provided here a concise overview of the technique, with a focus on its core principles. For a comprehensive exploration, references like [22] are highly recommended.

2.3.1 Proving non-negativity of a polynomial

To better understand how to construct such a hierarchy of SDPs that approximately solve a polynomial optimization problem, we can consider a simpler problem: demonstrating the non-negativity of a given polynomial:

$$\forall x_1, x_2, \cdots x_n : \quad p(x_1, \cdots x_n) \geq 0. \tag{2.9}$$

We use the notation POS to represent the set of all (globally) non-negative polynomials.

2.3.1.1 *Proving if a function is sum of squares*

A simple method to determine the non-negativity of a polynomial is to show that it can be expressed as a sum of squares. We define SOS_d as the set of all polynomials of degree less than or equal to d that have a sum of squares decomposition.

The problem of determining whether a given polynomial can be expressed as a sum of squares - i.e. $p \in \mathsf{SOS}_d$, can be cast as an SDP. To understand why, consider a polynomial $p \in \mathsf{SOS}_d$ of degree d. Assuming it has a sum of squares decomposition, $p = \sum_i p_i^2$, where each p_i is a polynomial of degree less than or equal to $d/2$. Each p_i can be expressed in matrix form:

$$p_i = [1, x_1, x_2, \cdots, x_1x_2, \cdots] \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{12} \\ \vdots \end{bmatrix},$$

i.e, $p_i = \mathbf{x}^T\mathbf{a}$, where $\mathbf{x}$ is a vector consisting of monomials and $\mathbf{a}$ is the vector corresponding coefficients of the monomial. Then, it is easy to see that

$$p_i^2 = \mathbf{x}^T\mathbf{a}^T\mathbf{a}\mathbf{x}.$$

Note that the matrix $\mathbf{a}^T\mathbf{a}$ is positive semi-definite. We can immediately deduce that if a polynomial $p \in \mathsf{SOS}_d$, then there exists a matrix $\mathbf{A}$ such that $\mathbf{A}$ is a sum of matrices $\mathbf{a}^T\mathbf{a}$, and is itself positive semi-definite. It is also evident from straightforward reasoning that if $p = \mathbf{x}^T\mathbf{A}\mathbf{x}$ for some $\mathbf{A} \succeq 0$, then p can be expressed as a sum of squares of polynomials. The simplest approach do prove this is to express $\mathbf{A}$ through its spectral decomposition to explicitly construct the sum-of-squares polynomials. Hence, a polynomial p belongs to the set SOS_d if and only if there exists a positive semi-definite matrix $\mathbf{A}$ such that $p = \mathbf{x}^T\mathbf{A}\mathbf{x}$.

Let's now consider a scenario where we have an arbitrary polynomial g, and our objective is to determine whether $g \in \mathsf{SOS}_d$. To achieve this, we aim find a suitable $D \succeq 0$ such that $g = \mathbf{x}^T D\mathbf{x}$.

Since g is a polynomial with a degree at most d, we can uniquely determine it through a finite set of constraints. These constraints, for instance can take the form $g(c_0^{(0)}, c_1^{(0)}, \cdots, c_d^{(0)}) = b_0$, $g(c_0^{(1)}, c_1^{(1)}, \cdots, c_d^{(1)}) = b_1$, and so forth. These constraints

can be expanded as linear conditions on the matrix D:

$$\mathbf{c}_i^T D \mathbf{c}_i = b_i.$$

Thus, if there is a solution to the following problem:

$$\begin{aligned} \max \quad & 0 \\ \text{s.t.} \quad & \forall i \in \{1, 2, \cdots, k\} : \mathbf{c}_i^T D \mathbf{c}_i = b_i \\ & D \succeq 0. \end{aligned} \tag{2.10}$$

Then we have successfully decomposed g into a sum of squares of polynomials. Conversely, if no solution exists for the above problem, we have demonstrated that g cannot be expressed as a sum of squares.
Optimization problems with trivial objective functions such as (2.10) are often used to prove the existence (or lack of existence) of solutions of simultaneous equations and inequalities. The solution to such an optimization problem is 0 iff there the set of constraints can be simultaneously satisfied. Having no solution to such an optimization problem is proof that the constraints can not be simultaneously satisfied.

2.3.1.2 Using SOS condition to determine positivity

In the previous subsection 2.3.1.1, we have seen an efficient method to determine if a given polynomial p can be expressed as a sum of squares. It is immediately clear that any polynomial that can be expressed as a sum of squares is non-negative - i.e. SOS $\subseteq$ POS.

However, the intriguing question arises: Is it possible that SOS $=$ POS, or is it the case that SOS $\subset$ POS? If the former holds true, then indeed, we would have already found an efficient way to determine the positivity of any polynomial.
The answer to this problem is shown to be unfortunately in the negative. However, it turns out that there is a very interesting relationship between the sets SOS and POS. It can be shown that

$$p \text{ is non-negative} \quad \iff \exists \text{ polynomial } q \text{ such that } qp \in \text{SOS}.$$

In fact, we can reduce the search of polynomials q to be of the form $q_n = (x_1^2 + \cdots + x_d^2)^n$. Note that now there is a possibility of coming up with a way to show if a polynomial is positive. Consider the set $\text{SOS}^{(n)}$ defined as

$$\{p : \quad (x_1^2 + x_2^2 + \cdots x_d^2)^n p \in \text{SOS}\}.$$

Then $p \in \mathsf{POS}$ implies that there exists $n \in \mathbb{N}$ such that $p \in \mathsf{SOS}^{(n)}$. Thus, there is a hierarchy of sets, which will eventually cover all the positive polynomials. Note that we have now a hierarchy of sets $\mathsf{SOS}^{(n)}$ that converge to the set of positive polynomials from inside

$$\mathsf{SOS} \subseteq \mathsf{SOS}^{(1)} \subseteq \mathsf{SOS}^{(2)} \subseteq \cdots = \mathsf{POS}.$$

See Figure 2.1 for a schematic illustration of the hierarchy. It is also possible to come up with a converging hierarchy of semi-definite programs that converge to the set POS from outside. This discussion is beyond the scope of the thesis.

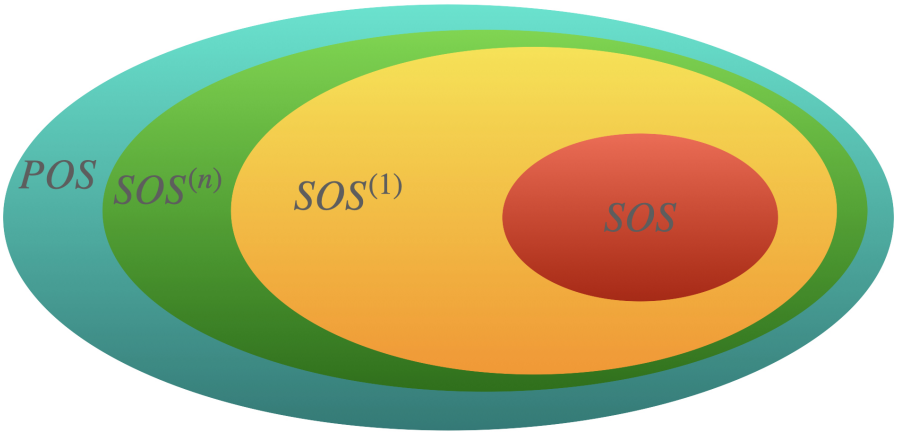


Figure 2.1: A schematic depiction of the converging sets $\mathsf{SOS}^{(n)}$ and the set POS.

2.3.2 SOLVING AN UNCONSTRAINED POLYNOMIAL OPTIMIZATION PROBLEM

Now that we possess a method for determining whether $p \in \mathsf{POS}$, we can also extend this same argument to address the unconstrained polynomial optimization problem

$$\min_{x \in \mathbb{R}^d} p(x),$$

by observing that it can be equivalently reformulated as the subsequent optimization problem:

$$\begin{aligned} \max_{\gamma \in \mathbb{R}} \quad & \gamma \\ \text{s.t.} \quad & p - \gamma \in \mathsf{POS}. \end{aligned} \tag{2.11}$$

This reformulation can be further transformed into a progressively convergent hierarchy of SDPs problems, expressed as follows:

$$\begin{aligned} \max_{\gamma\in\mathbb{R}} \quad & \gamma \\ \text{s.t.} \quad & p-\gamma \in \mathsf{SOS}^{(n)}. \end{aligned} \tag{2.12}$$

2.3.3 Solving a constrained polynomial optimization problem

We saw above that determining global non-negativity of a polynomial can be cast as a hierarchy of converging semi-definite programs. In order to solve the constrained polynomial optimization problem such as (2.8), we construct the following polynomial

$$f(x_1\cdots x_k) = p(x_1..x_n) + \sum_i \sigma_i(x_1..x_r)g_i(x_1..x_n) + \sum_j \rho_j(x_1\cdots x_l)h_j(x_1..x_n), \tag{2.13}$$

where $\sigma_i(x_1,\cdots x_r) \in \mathsf{POS}$ are known non-negative polynomials, and $\rho_j(x_1,\cdots x_r)$ are any polynomials. Note that $f(x_1,\cdots x_k)$ is non-negative whenever the constraints are satisfied - i.e. when $g_i(x_1,\cdots x_n) \geq 0$ and $h_j(x_1,\cdots x_n) = 0$. We can now find a set of solutions to our global polynomial optimization problems by solving the following problem

$$\begin{aligned} \max_{\gamma\in\mathbb{R}} \quad & \gamma \\ \text{s.t.} \quad & \exists f \text{ of the form (2.13)} \\ & f-\gamma \in \mathsf{POS}. \end{aligned} \tag{2.14}$$

We can again find suitable relaxations of this problem in terms of the semi-definite programs as discussed in the previous subsection.

These techniques are well known in the convex optimization literature. There are standard packages for many programming languages such as Python and Matlab that solve such optimization problems using techniques similar to the one discussed in this section. The package used to solve the optimization problems in this work is the NCPol2SDPA package [23]. Other semi-definite programs are solved using solvers PICOS [24].

2.4 Operational meaning of the min-entropy

In a randomness expansion protocol, the main objective is to quantify the amount of randomness produced by the protocol. The key question then arises: How can we quantify the amount of randomness? To address this query, we must accurately characterize the

broader situation in which we are operating.

Consider a secure laboratory where we have a random variable, denoted A. Intuitively, we would consider the random variable A as a source of randomness if its outcome cannot be accurately predicted. However, since we are dealing with a cryptographic scenario, we must be thorough and also account for the potential existence of an adversary for whom this random variable should also remain unpredictable. Thus, we will envision an imaginary adversary, whom we will call Eve or the Eavesdropper (or adversary), possessing certain information related to the random variable A. This information could include a correlated random variable or even a quantum state. However, Eve does not have direct access to the variable A.

In the most general case we allow that adversary to hold a quantum state ρ_E^a for every outcome a of the random variable A. This scenario can be effectively described using the Classical-Quantum (CQ) state:

$$\rho_{CQ} = \sum_a p_A(a) \, |a\rangle\langle a| \otimes \rho_E^a.$$

Now, imagine Eve attempts to determine the value of the random variable A. To achieve this, she performs a POVM, denoted as $\{\mathcal{F}_a\}_a$, on her state, where each outcome corresponds to a possible value of A. The probability that Eve correctly guesses the outcome is expressed as:

$$p_{\text{guess}} = \sum_a p_A(a) \mathcal{F}_a(\rho_E^a).$$

Given that the choice of measurement used by Eve can be arbitrary, it is reasonable to assume that she would select the optimal measurement strategy to maximize her guessing probability. With this in mind, quantifying the amount of randomness in a cryptographic scenario is related an optimization problem, which can be formulated as follows:

$$\begin{aligned} p_{\text{guess}}^* = \quad & \sup \sum_a p_A(a) \mathcal{F}_a(\rho_E^a) \\ \text{s.t.} \quad & \{\mathcal{F}_a\}_a \text{ is a POVM.} \end{aligned} \tag{2.15}$$

It is easy to see that the optimization problem described above can be written in terms of an SDP. This is due to the linearity of the objective function with respect to the variables $\{\mathcal{F}_a\}_a$, and $\{\mathcal{F}_a\}_a$ being effects can be represented in terms positive semi-definite operators $\{F_a\}_a$ by the relation $\mathcal{F}_a(.) = \text{tr}(F_a(.))$. Furthermore, the constraints that $\sum_a \mathcal{F}_a = \mathcal{I}$ (where $\mathcal{I}$ is the identity map) is a linear constraint $\sum_a F_a = \mathbb{1}$.

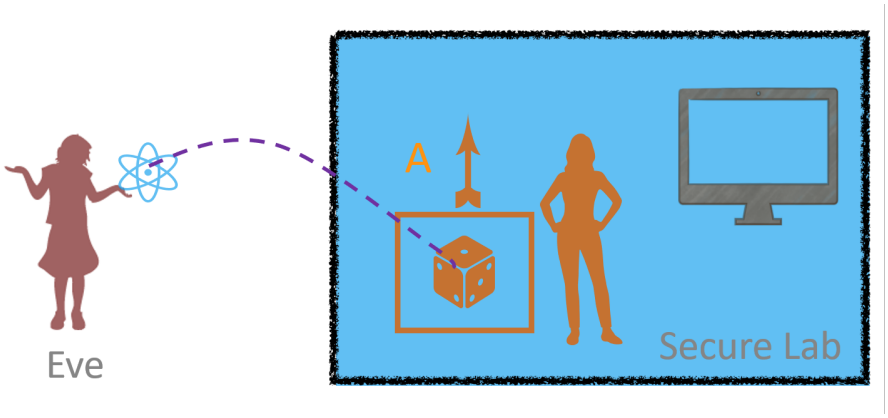


Figure 2.2: A diagram depicting the scenario in which an adversary is trying to guess the random variable A in the secure lab.

Intuitively, the amount of randomness should be a function of this optimal guessing probability p^*_{guess}. Formally, in quantum information theory, the amount of randomness in a protocol is quantified using the min-entropy [25]:

Definition 7 (Min-entropy)**.** The min-entropy of a state $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ is given by

$$H_{\min}(A|B)_{\rho_{AB}} \quad = \quad -\inf_{\sigma_B} \quad \{\lambda| \quad \rho_{AB} \preceq 2^{\lambda}(\mathbb{1}_A \otimes \sigma_B)\}, \tag{2.16}$$

where $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$.

To understand the significance of why min-entropy is an appropriate measure for randomness in a cryptographic scenario, we must understand that the primary objective of such a protocol is to generate random bits that are uniformly (or almost uniformly) distributed. Moreover, these bits should be independent and uncorrelated with any side information held by an adversary. Hence, at the end of the protocol, we require a string of uniformly distributed random numbers $\mathbf{R}$ uncorrelated with any system held by a potential adversary. Randomness extractors are procedures that "extract" such a random string $\mathbf{R}$ from a random-variable A (obtained as an output of a protocol) that could, in general, be correlated with the side-information of the adversary. Roughly speaking, a quantum-proof strong extractor Ext is a function that takes a random variable A and an input random seed U, outputting the desired string: $\mathbf{R} = \mathrm{Ext}(A, U)$. Technical results, like the generalized left over-hashing lemma [26], show that it is possible to extract approximately $H_{\min}(A|E)_{\rho_{AE}}$ such uniformly random bits from the random variable A given a particular classical-quantum state ρ_{AE} (that describes the protocol). Further details on randomness extractors are beyond the scope of the thesis.

However, as it stands, the definition of min-entropy is not very insightful or intuitive. We sketch the proof from [25] to show that the min-entropy and the guessing probability are in-fact related to each other.

To begin, let us introduce a new variable $\tilde{\sigma}_B = 2^{\lambda}\sigma_B$. Exploiting the monotonicity property of the logarithms, it becomes apparent that the min-entropy is given by $-\log_2(g)$, where g can be computed through the following optimization problem:

$$\begin{aligned} g = \quad & \inf \operatorname{tr}(\tilde{\sigma}_B) \\ \text{s.t.} \quad & (\mathbb{1}_A \otimes \tilde{\sigma}_B) - \rho_{AB} \succeq 0 \\ & \tilde{\sigma}_B \succeq 0. \end{aligned} \tag{2.17}$$

We will now show that this function g is , in fact, the best guessing probability p^*_{guess}. Using the duality theory of optimization [27], it is known that for every SDP, a corresponding dual SDP can be constructed, and the solutions to both problems are identical. In the case of the problem above, the dual for the SDP above is:

$$\begin{aligned} & \sup \operatorname{tr}(Y\rho_{AB}) \\ \text{s.t.} \quad & Y \succeq 0 \\ & \operatorname{tr}_A(Y) = 1. \end{aligned} \tag{2.18}$$

We can now use the Choi–Jamiokowski isomorphism [28] to relate the non-negative operator Y in terms of a channel $\mathcal{E}$ as

$$Y = d_A(\mathbb{1}_A \otimes \mathcal{E})(|\Phi_{AB}\rangle\langle\Phi_{AB}|), \tag{2.19}$$

where $|\Phi_{AB}\rangle = \sum_x \frac{1}{d_A} |x,x\rangle$ is the maximally entangled state and $d_A = \dim(\mathcal{H}_A)$. As $\operatorname{tr}(\mathbb{1}_A \otimes \mathcal{E}(|\Phi_{AB}\rangle\langle\Phi_{AB}|)\rho_{AB}) = \operatorname{tr}\Big(|\Phi_{AB}\rangle\langle\Phi_{AB}|\, \mathbb{1}_A \otimes \mathcal{E}^\dagger(\rho_{AB})\Big)$, we can re-write the trace $\operatorname{tr}(Y\rho_{AB})$ as:

$$\operatorname{tr}\Big(|\Phi_{AB}\rangle\langle\Phi_{AB}|\, \mathcal{I} \otimes \mathcal{E}^\dagger(\rho_{AB})\Big) = \sum_x \langle x,x| \left(\mathbb{1} \otimes \mathcal{E}^\dagger(\rho_{AB})\right) |x,x\rangle \,. \tag{2.20}$$

This gives the optimization problem,

$$\begin{aligned} g = \quad & \sup \sum_x \langle x,x| \left(\mathbb{1} \otimes \mathcal{F}(\rho_{AB})\right) |x,x\rangle \\ \text{s.t.} \quad & \mathcal{F} \text{ is a quantum channel,} \end{aligned} \tag{2.21}$$

where $\mathcal{F} = \mathcal{E}^\dagger$ is any channel. We return to the case of computing the min-entropy for the CQ states of the form

$$\rho_{AB} = \sum_a p_A(a) \, |a\rangle\langle a| \otimes \rho^a_B. \tag{2.22}$$

Substituting the explicit form of ρ_{AB} in the objective function of eqn. (2.21) gives the following objective function:

$$\sum_a p_A(a) \langle a| \mathcal{F}(\rho_B^a) |a\rangle \equiv p_{\text{guess}}. \tag{2.23}$$

The collection of maps $\{\mathcal{F}_a \equiv \langle a| \mathcal{F}(.) |a\rangle\}_a$ without any loss of generality is an arbitrary POVM, due to the fact that

$$\forall \rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) : \qquad \sum_a \langle a| \mathcal{F}(\rho) |a\rangle = \text{tr}(\mathcal{F}\rho_{AB}) = \text{tr}(\rho). \tag{2.24}$$

Note that above we have used the fact that $\mathcal{F} = \mathcal{E}^\dagger$ can be taken to be trace-preserving. For consistency, unless stated otherwise, throughout the thesis, we have reserved the system E for the system of the Eavesdropper or the adversary.

2.5 QUANTIFYING RANDOMNESS IN A PROTOCOL

The framework described in the previous section is very general and can be useful for any cryptographic scenario. In this thesis, our primary interest is to quantify randomness in randomness expansion protocols, which has more structure. Randomness expansion protocols are inherently sequential by design; that is, these protocols are typically carried out over multiple rounds, each identified by $i \in \{1, \cdots, n\}$. In each round, a sub-protocol is performed, and this process is repeated n times. For instance, in the Device Independent scenario, the sub-protocol might correspond to a single CHSH test. The randomness expansion protocols, in general, may be described using an initial state shared by the lab and Eve along with a sequence of channels, each representing a sub-protocol as shown in Figure 2.3.

In the protocols we consider in this thesis, each round $i \in \{1, \cdots, n\}$ requires generating a random number D_i from an an existing source of randomness. By the end of round i, the protocol produces another random number, C_i. When considering CHSH Device Independent protocols, $D_i \equiv X_i, Y_i$ are the inputs for Alice and Bob in each round, and $C_i \equiv A_i B_i$ represent the outputs of a round of the CHSH test (see Section 3.2 for more details). Each round is characterized by a channel $\mathcal{N}_i$, and the overall protocol is defined by the collection of channels $\{\mathcal{N}_i\}_i$ where i runs from 1 to n.

The protocol starts with the initial state $\rho^0_{R_A E}$, where R_A represents the laboratory system, and E is the system in possession of Eve. In randomness expansion protocols,

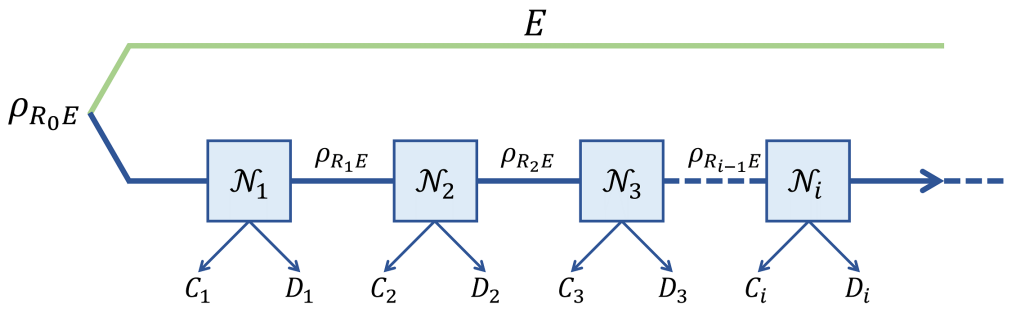


Figure 2.3: A protocol of randomness expansion in terms of sub-protocols. ρ_{R_0E} is the initial state of the protocol and EAT channels.

this is often an unknown, pre-programmed quantum resource. In each round, the channel $\mathcal{N}_i$ acts on the state $\rho^{i-1}_{R_{i-1}E}$ to output classical random variables C_i, D_i and a state $\rho^{i}_{R_i,E}$ for the next round[2]. For CHSH tests, R_0 denotes the initial pre-shared quantum state between Alice, Bob, and Eve. We further make the assumption that inputs D_i are independent random variables, and are uncorrelated with any other input random variables for any round other than round i. Moreover, D_i cannot be correlated with any outputs $A_1, A_2, \cdots, A_{i-1}$ generated before round i.

After the protocol is finished, the objective is to determine the randomness in the collection $\mathbf{C} = (C_1, C_2, \cdots, C_n)$ (or $\mathbf{AB} = (A_0, B_0, \cdots, A_n, B_n)$ in the DI protocol). This calculation is done conditioned upon a specific event, Ω, taking place. In our context, Ω is the event that protocol does not abort. Recall that, a DIRNE protocol aborts if the CHSH score ω does not exceed a given threshold score. In general, the event Ω (abort condition) is determined by the input-output statistics $\{C_i, D_i\}_i$. Note that computing the randomness conditioned only on the input and outputs is in the spirit of Device Independence. The input-output statistics can be directly observed in the laboratory, and therefore are known quantities to the party conducting the protocol. The internal mechanism of the devices used to generate the experimentally observed statistics is dictated by the pre-shared initial state $\rho_{R_0,E}$ and the channels $\{\mathcal{N}_i\}_i$ of the protocol. As the state and the channels are treated as unknown in the protocol, we make no assumption on the inner workings of the devices other than the fact that ρ_{R_0E} is a valid state in Quantum Theory and $\{\mathcal{N}_i\}_i$ are valid quantum channels.

We say that the protocol generates randomness if output string $\mathbf{C}$ contains secure randomness; that is, given access to the system E and the input string $\mathbf{D} = (D_1, D_2, \cdots, D_n)$,

[2]In reality, devices may receive new states each round. However, it is equally valid to assume that the devices have pre-shared entangled quantum resources needed for the entire protocol.

the output string $\mathbf{C} = (C_1, C_2, \cdots, C_n)$ cannot be determined with certainty. Consequently, using the discussion in the previous section, our main aim is to determine $H_{\min}(\mathbf{C}|\mathbf{D}E)_\rho$ in a protocol that does not abort.

This calculation needs to be done considering the worst-case scenario, where ρ_{R_0E} is unknown to the party executing the protocol but fully accessible to Eve. Moreover, the channels $\mathcal{N}_i$ are also known to Eve. The main assumption here is that Eve cannot access the laboratory system after the protocol begins. Given this immense power granted to Eve, tackling this optimization problem is extremely challenging. Therefore instead of finding the exact min-entropy in a given protocol, it is acceptable to determine a lower bound on the min-entropy. This lower bound tells us the minimum randomness we can "safely" extract from the protocol. However, it is crucial that this lower bound is not too far off from the actual value or else we will waste randomness.

In the literature, incremental progress was made towards getting lower bounds on the min-entropy for a cryptographic protocol. It began with the asymptotic equipartition theorem [29], then advanced with the Entropy Accumulation Theorem (EAT), and subsequent proofs catering to more generalized settings [30]. In this thesis, we use the EAT to compute randomness in a protocol. Informally, the EAT states that

$$H_{\min}(\mathbf{C}|\mathbf{D}E)_\rho \geq n \inf H(C|DE) - \sqrt{n}v, \tag{2.25}$$

where $H(C|DE)$ is the single round von Neumann entropy. Here $\inf$ is taken over all single-round strategies, which would reproduce the observed statistics if executed in an i.i.d. manner.

EAT significantly simplifies the challenge of determining the extractable randomness for the protocol. Without EAT, one would need to account for every individual and potentially undefined channel $\mathcal{N}_i$ and any arbitrary initial state ρ_{R_0E}, whereas EAT permits us to focus solely on the von Neumann entropy of a representative single round of the protocol. Moreover, as $n \to \infty$, this lower bound becomes tight, implying that we can use the quantity $\inf H(C|DE)$ as the asymptotic randomness rate (amount of extractable randomness per round) in a protocol. The error term $\sqrt{n}v$ is intricate and defined via the min-tradeoff function. These technical details will be elaborated on in the next section of this thesis.

However, the EAT does not fully resolve the problem of determining rates, leaving the crucial task of optimizing single-round von Neumann entropies. This optimization becomes extremely crucial when developing these protocols and is the central theme of many chapters in this thesis.

2.6 ENTROPY ACCUMULATION THEOREM

In this section we state the Entropy Accumulation Theorem (EAT) more formally with all relevant details. The theorem is phrased in terms of a set of channels $\{\mathcal{M}_i\}_i$ called EAT channels, where $\mathcal{M}_i : \mathcal{S}(R_{i-1}) \to \mathcal{S}(C_i D_i U_i R_i)$ (here $i \in \{1, 2, \cdots, n\}$).

Definition 8 (EAT Channels)**.** Let $\{R_i\}_{i=0}^n$ be arbitrary quantum systems and $\{C_i\}_{i=1}^n$, $\{D_i\}_{i=1}^n$, and $\{U_i\}_{i=1}^n$ be finite dimensional classical systems. Suppose that U_i is a deterministic function of C_i, D_i and that $\{\mathcal{M}_i\}_{i=1}^n$, $\mathcal{M}_i : \mathcal{S}(R_{i-1}) \to \mathcal{S}(C_i D_i U_i R_i)$ are a set of quantum channels. These channels form a set of EAT channels if for all $\rho_{R_0 E} \in \mathcal{S}(R_0 E)$ the state $\rho_{\mathbf{CDU}R_n E} = (\mathcal{M}_n \circ \ldots \mathcal{M}_1)(\rho_{R_0 E})$ after applying the channels satisfies $I(C_1^{i-1} : D_i | D_1^{i-1} E) = 0$, where I is the mutual information, and C_1^{i-1} is shorthand for $C_1 C_2 \ldots C_{i-1}$.

In the context of protocols, the register U_i records the score for the round i. Each EAT channel for the randomness expansion protocols is a set of maps of the form

$$\mathcal{M}_i(\rho) = \sum_{c,d} |c\rangle\langle c| \otimes |d\rangle\langle d| \otimes |u(c,d)\rangle\langle u(c,d)| \otimes \mathcal{M}_i^{c,d}(\rho)\,, \tag{2.26}$$

where $u(c,d)$ records the score in the protocol, and each $\mathcal{M}_i^{c,d}$ is a subnormalized quantum channel from $\mathcal{S}(R_{i-1})$ to $\mathcal{S}(R_i)$. The joint distribution of the classical variables C_i and D_i is

$$p_{C_i D_i}(c,d) := \mathrm{tr}\Big(\mathcal{M}_i^{c,d}(\rho)\Big)\,. \tag{2.27}$$

Definition 9 (Frequency distribution function)**.** Let $\mathbf{U} = U_1 U_2 \ldots U_n$ be a string of variables. The associated frequency distribution is

$$\mathrm{Freq}_{\mathbf{U}}(u) := \frac{|\{i \in \{1, \ldots, n\} : U_i = u\}|}{n}\,. \tag{2.28}$$

In the above equation (eqn. 2.28), the notation $|.|$ is used to represent the cardinality of a set. We use this notation to denote the cardinality of sets throughout this thesis.

Definition 10. Given a set of channels $\mathfrak{G}$ whose outputs have a register U, the set of achievable score distributions is

$$\mathcal{Q}_{\mathfrak{G}} := \{p_U : \mathcal{M}(\rho)_U = \sum_u p_U(u)\,|u\rangle\langle u| \text{ for some } \mathcal{M} \in \mathfrak{G}\}. \tag{2.29}$$

For the spot checking protocol (introduced in Chapter 6), there is an additional quantity of interest

$$\mathcal{Q}_{\mathfrak{G}}^{\gamma} := \{p_U : p_U(\perp) = (1-\gamma) \text{ and } p_U(u) = \gamma \tilde{p}_U(u) \text{ with } \tilde{p}_U \in \mathcal{Q}_{\mathfrak{G}}\}. \tag{2.30}$$

The above definition is given for completeness here. The variable $\perp$ is defined in Protocol 1 (see chapter 6 for further details).

Definition 11 (Rate function)**.** Let $\mathfrak{G}$ be a set of EAT channels. A *rate function* $\mathrm{rate} : \mathcal{Q}_{\mathfrak{G}} \to \mathbb{R}$ is any function that satisfies

$$\mathrm{rate}(q) \leq \inf_{(\mathcal{M},\rho_{RE})\in\Gamma_{\mathfrak{G}}(q)} H(C|DE)_{(\mathcal{M}\otimes\mathcal{I}_E)(\rho_{RE})}\,, \tag{2.31}$$

where

$$\Gamma_{\mathfrak{G}}(q) := \{(\mathcal{M},\rho_{RE}) : (\mathcal{M}\otimes\mathcal{I}_E)(\rho_{RE})_U = \sum_u q(u)\,|u\rangle\langle u| \text{ for some } \mathcal{M}\in\mathfrak{G}\} \tag{2.32}$$

is the set of states and channels that can achieve distribution q.

Definition 12 (Min-tradeoff function)**.** A function $f : \mathcal{Q}_{\mathfrak{G}} \to \mathbb{R}$ is a *min-tradeoff function* if f is an affine rate function. Since min-tradeoff functions are affine, we can naturally extend their domain to all probability distributions on U, denoted $\mathcal{P}$.

The entropy accumulation theorem then can be stated as follows (this is Theorem 2 of [31], which is a generalization of the results of [32]).

Theorem 1. *Let f be a min-tradeoff function for a set of EAT channels $\mathfrak{G} = \{\mathcal{M}_i\}_i$ and $\rho_{\mathbf{CDU}E}$ be the output after applying these channels to initial state ρ_{RE}. In addition let $\epsilon_h \in (0,1)$, $\alpha \in (1,2)$ and $r \in \mathbb{R}$ and Ω be an event on $\mathbf{U}$ that implies $f(\mathrm{Freq}_{\mathbf{U}}) \geq r$. We have*

$$H^{\epsilon_h}_{\min}(\mathbf{C}|\mathbf{D}E)_{\rho_{\mathbf{CD}E|\Omega}} > nr - \frac{\alpha}{\alpha-1}\log\left(\frac{1}{p_\Omega(1-\sqrt{1-\epsilon_h^2})}\right) + n\inf_{p\in\mathcal{Q}_{\mathfrak{G}}}\left(\Delta(f,p) - (\alpha-1)V(f,p) - (\alpha-1)^2K_\alpha(f)\right)\,, \tag{2.33}$$

where $\Delta(f,p) = \mathrm{rate}(p) - f(p)$, and

$$V(f,p) = \frac{\ln 2}{2}\left(\log(1+2d_C) + \sqrt{2+\mathrm{Var}_p(f)}\right)^2$$

$$K_\alpha(f) = \frac{1}{6(2-\alpha)^3\ln 2}2^{(\alpha-1)(\log(d_C)+\mathrm{Max}(f)-\mathrm{Min}_{\mathcal{Q}_{\mathfrak{G}}}(f))}\ln^3\left(2^{\log(d_C)+\mathrm{Max}(f)-\mathrm{Min}_{\mathcal{Q}_{\mathfrak{G}}}(f)} + \mathrm{e}^2\right)\,,$$

and we have also used

$$\mathrm{Max}(f) = \max_{p\in\mathcal{P}} f(p)$$

$$\mathrm{Min}_{\mathcal{Q}_{\mathfrak{G}}}(f) = \inf_{p\in\mathcal{Q}_{\mathfrak{G}}} f(p)$$

$$\mathrm{Var}_p(f) = \sum_u p(u)\left(f(\delta_u) - \mathbb{E}(f(\delta_u))\right)^2\,,$$

and δ_u is the deterministic distribution with outcome u.

To use this theorem we have to assign the variables C_i and D_i to the parameters in the protocol. For example, as stated in the previous section, in the context of CHSH-based Device Independent Protocols, the variables C_i can be output variables A_iB_i and the variable D_i can be taken to be the input variables X_iY_i.

Part I

Device Independent Protocols

— 3 —

Introduction to Device Independent Protocols

3.1 INTRODUCTION

Bell's theorem states that quantum mechanics is not compatible with local hidden variable theories. A Bell test is a non-local experiment that consists of two (or more) space-like separated (or non-communicating) observers who share an entangled quantum state. At the beginning of the experiment, each observer generates a discrete random input (with a finite number of possible inputs). Based on this input, these observers perform a measurement on their shared state to generate an output. The observers repeat this process for a large number of rounds to get some input-output statistics. A Bell inequality is a relation on the joint input-output statistics of both these parties that is satisfied if these statistics can be obtained by a local-hidden variable theory. Therefore, if the obtained input-output statistics indicate the violation of a Bell inequality, then those statistics cannot have arisen from a local deterministic behaviour.

In the context of Device Independent randomness expansion (DIRNE), the main idea is that the ability to violate a Bell inequality implies that the devices doing so must be generating randomness [1, 2] (see section 3.2 for further details). Thus, in a sense, the protocol self-tests[1] [33] the devices during its operation, leading to enhanced security. Although challenging to accomplish, recently the first experimental demonstrations of DIRNE were performed [8, 31, 34], following earlier experiments considering randomness generation [3, 7, 35].

On the theoretical side, the main difficulty is then to calculate how much extractable randomness there is as a function of the Bell violation. This task is made more challenging

[1] Roughly speaking, a protocol is called a self testing protocol if we are able to infer the underlying physical process solely from the observable outcomes of the protocol.

by the lack of structure of the problem: no assumption is made on how the devices operate and so one has to account for arbitrary pre-shared entanglement, arbitrary measurements, and strategies that may be adaptive between the rounds. An increasingly sophisticated series of proofs [36–38] leads to two techniques for dealing with this exist in the literature: the quantum probability estimation framework [13] and the entropy accumulation theorem (EAT) [14, 15] (see Section 2.6 for details). We use the latter in this work. Informally speaking, the EAT states that the amount of extractable randomness in the full string of outputs is to leading order n times the von Neumann entropy of a single-round strategy that would give the observed score if used in an i.i.d. way. In other words, the EAT implies that if we can solve the problem for an i.i.d. adversary, then we can get a bound for the general case.

In a DIRNE protocol, randomly chosen inputs are made to two separated devices so that each device cannot determine the input of the other device. We use X and Y to label the inputs, and A and B to label the outputs, taking into account an adversary with side information E. This side information could be quantum; the general strategy allows for the adversary holding the E part of a state $\rho_{A'B'E}$, with the A' and B' systems retained by the devices. Each input X to the first device corresponds to a measurement on A' yielding outcome A; similarly, each input Y to the other device corresponds to a measurement on B' resulting in outcome B. Two quantities of interest arise, both dependent on the post-measurement state: the first is the score in a non-local game – a function of the conditional distribution $p_{AB|XY}$and the second, the von Neumann entropy of either one or both of the outputs. Specifically, we aim to express the minimum von Neumann entropy in terms of the score. In this work, we study six von Neumann entropies: $H(AB|X=0,Y=0,E)$, $H(AB|XYE)$, $H(AB|E)$, $H(A|X=0,Y=0,E)$, $H(A|XYE)$, and $H(A|E)$ [2].

In essence, using the EAT, the problem of computing extractible randomness is reduced to finding the smallest von Neumann entropy of the outputs conditioned on the adversary's side information and the inputs, with the property that the state and measurements used would give a particular Bell value. More precisely, a strategy (for the adversary) corresponds to picking a quantum state $\rho_{A'B'E}$ and for each possible input $X=x$ a POVM $\{M_{a|x}\}_a$ on A', and for each $Y=y$ a POVM $\{N_{b|y}\}_b$ on B'. From this, the

[2] In the one-sided cases, conditioning on the variable Y is unnecessary, but it is retained for notational symmetry.

implementation of the protocol corresponds to repeated actions of the channel:

$$\begin{aligned} \mathcal{N} \quad &: \quad \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_X \otimes \mathcal{H}_Y) : \\ \sigma \quad &\mapsto \quad \sum_{abxy} p_{XY}(x,y) \, |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes |x\rangle\langle x| \otimes |y\rangle\langle y| \, \mathrm{tr}\left((M_{a|x} \otimes N_{b|y})\sigma\right), \end{aligned}$$

where p_{XY} is the input distribution used in the protocol. Applying this channel to $A'B'$ and acting as identity on E generates the final state $\tau_{ABXYE} = (\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$. The Bell value is a linear function on this state, and we seek to minimize an entropy (e.g., $H(AB|XYE)$) for this state, over all strategies that have a given Bell value (see later for a discussion of different entropies).

While the minimization of the single round entropic quantities above proves simpler than the direct optimization of the min-entropy, this minimization remains challenging. The main challenge is that the von Neumann entropies are non-linear. Additionally, there is no preliminary upper bound on the dimensions of the systems A', B', and E. For instance, some Bell inequalities suggest that the maximum quantum violation cannot be realized if A' and B' are finite dimensional [39]. Moreover, evidence suggests this remains true even when X and Y are binary, and A and B have only three possible outcomes [40]. However, when A, B, X, and Y are all binary, Jordan's lemma [41] asserts that there is no loss in generality when considering a convex combination of strategies wherein A' and B' are two-dimensional. This observation paved the way for establishing a tight lower bound on the one-sided entropy based on the CHSH score [42] and is pivotal for this study.

In Chapter 4, we discuss computing the von Neumann entropy bounds, later presenting numerically generated upper bound curves for each of the six quantities for the protocols based on violation of the CHSH inequality 4.12. By employing Jordan's lemma and other technical maneuvers [42], we can reduce the problem to seven real parameters (three for state specification and four for measurement selection), making it suitable for numerical optimization. Using heuristic numerical techniques we determined bounds which serve as useful references for the optimal values von Neumann entropic quantities for the six different cases as a function of the CHSH score, ω (notably, an analytic bound exists for the first case [42]). For $H(A|XYE)$ and $H(AB|XYE)$, we also propose conjectured analytic forms for the curves.

In Chapter 5, rigorous lower bounds on the entropic quantities $H(A|XYE)$ and $H(AB|X = 0, Y = 0, E)$ as a function of the CHSH score, ω, are also derived. We show that lower bounds for these quantities can be calculated by solving an optimization problem over

three real parameters. This realization lets us partition the domain of these new optimization problems into cuboids and compute the objective function on their edges. The objective function's value within each cuboid is underestimated using Taylor's theorem, allowing us to obtain a lower bound on the entropies by taking the minimum over all the function's under-estimations on each cube.

Using a similar method of partitioning the domain into (hyper) cuboids, combined with techniques for lower bounding polynomial optimization problems, we have also found reliable lower bounds for the entropy $H(AB|XYE)$.

The derived entropy lower bounds can be made arbitrarily tight by refining the partition, albeit with increased computation time (see Chapter 5). If one needs to compute the min-tradeoff function for application to the EAT, a more refined partition can be used since entropy only needs computation for a limited CHSH score range. Thus, our methodology has relevance in practical applications of DIRNE protocols. For the entropy $H(AB|X = 0, Y = 0, E)$, our findings are compared with the recent breakthrough technique [43] for optimizing von Neumann entropies. We find that our the lower bounds are very close to the numerical upper bounds for both $H(A|XYE)$ and $H(AB|X = 0, Y = 0, E)$. For the case $H(AB|X = 0, Y = 0, E)$, our lower bounds surpass the one ones obtained using the technique in [43]. Consequently, we conjecture that the numerical bounds established for all six entropic quantities are tight. Hence, in our work, the conjectured bounds are used to determine randomness rates across various protocols for randomness expansion.

Given the challenges associated with optimizing von Neumann entropies in randomness expansion protocols, the one-sided quantity $H(A|X = 0, Y = 0, E)$ has often been used due to its existing analytic bound [42]. However, as this omits Bob's output, it can be deemed wasteful in terms of generating randomness. With our new bounds available, we can now employ the corresponding two-sided quantities in randomness expansion protocols, such as the spot-checking protocol (refer to Chapter 6). This bound also facilitates the calculation of randomness expansion in protocols using (heavily) biased input randomness, allowing for the closure of the locality loophole present in the spot-checking protocol. Furthermore, the lower bounds for the entropy $H(AB|XYE)$ allow for the recycling of input randomness, ensuring a more efficient utilization of all resources.

Our newly derived entropies are employed in a protocol for Device Independent randomness generation, allowing two key improvements over prior works:

- The protocol exploits two-sided randomness.
- It recycles input randomness.

These modifications not only improve the randomness rate but also close the locality loophole associated with the spot-checking protocol. In Chapter 6, we outline rate curves with finite round numbers (employing the EAT with the updated numerical bounds on the single-round von Neumann entropy). As anticipated, the gains from the i.i.d. case transition seamlessly to the finite regime. For example, when taking the experimental conditions from [31], using two-sided randomness coupled with randomness recycling culminates in a rate increase of multiple orders of magnitude.

3.2 BELL'S THEOREM AND RANDOMNESS

Before delving into the intricacies of DI protocols for randomness expansion, we will first discuss the Bell's theorem and its role in the randomness expansion protocols.
As depicted in Figure 3.1, the simplest scenario of the Bell setup involves two spatially separated *or non communicating* parties, say Alice and Bob, sharing a common resource. This resource could be, in general, a quantum state.
A Bell test is conducted over several rounds. In each round, the procedure remains the same: an input X is chosen randomly and dispatched to Alice's device, and similarly, an input Y is selected and sent to Bob. When Alice and Bob conduct measurements based on their randomly chosen inputs X and Y, their outcomes are denoted as A and B, respectively. After numerous repetitions, we can determine the conditional probability distribution $p_{AB|XY}$.
John Bell posed the question: can this probability distribution be derived "classically"[3] – in other words, can the outputs A, B be solely predetermined by the input random variables X, Y, and a classical random variable Λ that could, in theory, be known to any entity within the past light cone of both Alice and Bob?
For this scenario, we consider the simplest situation where the inputs X, Y and the outputs A, B are binary. The CHSH score, represented as ω, can then be defined as:

$$\omega := \frac{1}{4}\left(\sum_a p_{AB|00}(a,a) + \sum_a p_{AB|01}(a,a) + \sum_a p_{AB|10}(a,a) + \sum_a p_{AB|11}(a,a\oplus 1)\right), \tag{3.1}$$

[3]Here the word "classical" is used to refer to any theory which obeys local determinism.

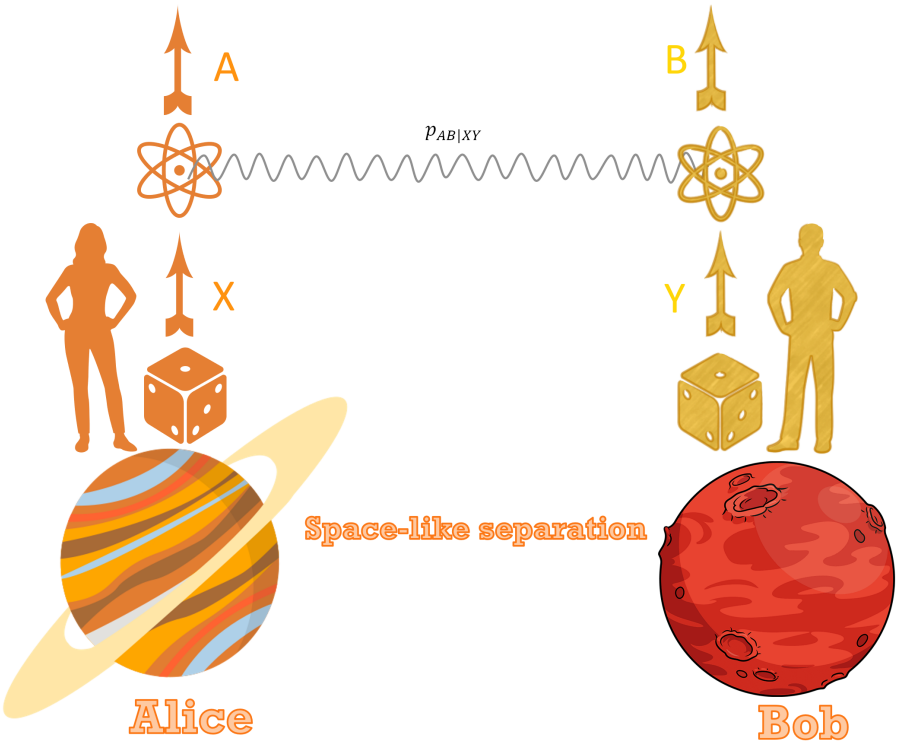


Figure 3.1: A diagram depicting a typical Bell scenario involving two parties.

which can be determined solely using collected input and output statistics $p_{AB|XY}$. Bell's theorem states that if $\omega > \frac{3}{4}$, the outputs must have been generated by a non-classical resource. This means that the outcomes of the given measurement cannot be achieved in a local deterministic fashion. Thus they cannot be determined with certainty by anyone including any adversary, who has the full knowledge of the inner workings of the device. This principle lays the groundwork for the generation of random numbers within the Bell type setup.

Bell's theorem is often also described using the scenario where two non-communicating parties – Alice and Bob are playing a non-local game. In each round, they are posed a question represented by the value of the input of random variables $X \in \{0, 1\}$ and $Y \in \{0, 1\}$. They are then expected to output $A, B \in \{0, 1\}$. Alice and Bob are considered to have won a round of the game if $XY = A \oplus B$, and they lose if this condition is not met. If Alice and Bob play this game (referred to as the CHSH game) using local deterministic strategies, their maximum winning probability is $\leq \frac{3}{4}$. If they wish to win with a probability higher than this, they must employ a non-classical strategy. With the understanding that the outputs A, B are random given the random inputs X, Y

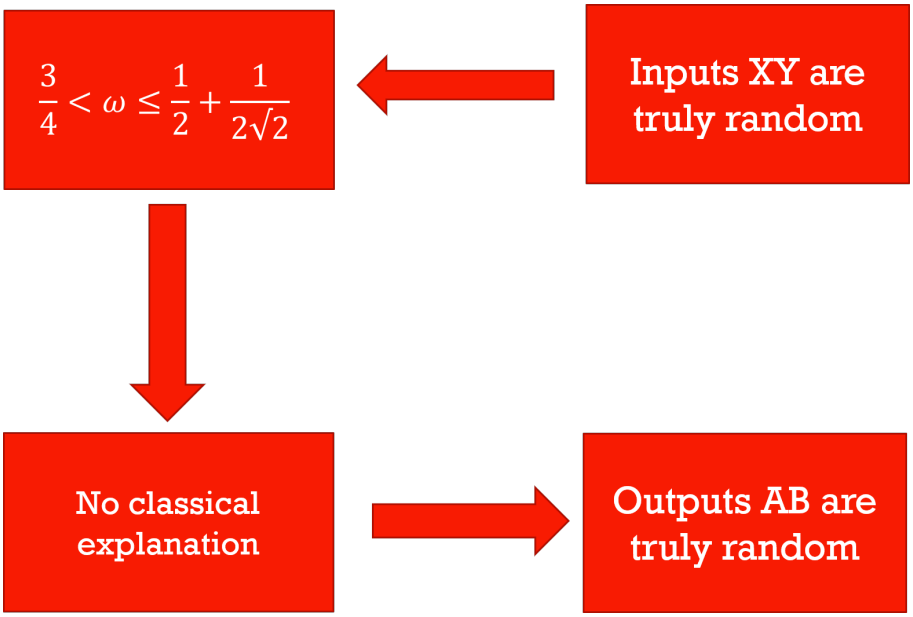


Figure 3.2: A diagram depicting an informal summary of Bell's theorem.

and a CHSH score $\omega \geq \frac{3}{4}$, the next question is to consider how much randomness can be extracted from the outputs. As discussed in Chapter 2, the randomness per CHSH test is roughly characterized by the von Neumann entropy $\inf H(AB|XYE)$. It is important to note that this value represents the randomness per round in the asymptotic limit (i.e., the randomness when the CHSH test is repeated indefinitely). The infimum is taken over all strategies that achieve a CHSH score ω, assuming the strategy is performed in an i.i.d. manner (see Chapter 4 and 5 for details). Note that other entropic quantities such as $\inf H(AB|X=0,Y=0,E)$ may also be of interest depending upon the protocol being performed. In the next section, we discuss the role of different entropic quantities that are useful in Device Independent protocols based on the violation of the (generalized) CHSH inequality.

In the literature, the inequality $\omega > 3/4$ is called a CHSH inequality. In a more general setting, if $p_{AB\cdots|XY\cdots}$ is a valid n partite probability distribution, a Bell inequality can be defined as $\mathcal{B}(p_{AB\cdots|XY\cdots}) > l$, where $\mathcal{B}$ is a linear functional on the probability distribution $p_{AB\cdots|XY\cdots}$ and l is the maximum value of $\mathcal{B}(p_{AB\cdots|XY\cdots})$ that can be achieved by a probability distribution corresponding to any local deterministic strategy.

It is worth highlighting that the CHSH inequality is not the sole Bell inequality, even when considering the two-input, two-output, two-party scenario. Distinct classes of Bell inequalities exist for this simplest scenario, such as the tilted Bell inequalities [44, 45].

Another category pertains to optimal inequalities tailored for randomness expansion [46]. Yet, the CHSH inequality violation remains the most extensively researched in context of DIRNE protocols. The collection of CHSH inequalities, defined by permutations of inputs X, Y and outputs A, B in the CHSH score above, stands out as, in the vector space of all probability distributions, the CHSH inequalities serve as facets of the polytope formed by all the local deterministic distributions [47]. In this dissertation, while determining randomness rates for our protocols, we employ generalized Bell inequalities expressed as:

$$\omega := \sum_a \Big(\gamma_{00} p_{AB|00}(a,a) + \gamma_{01} p_{AB|01}(a,a) + \gamma_{10} p_{AB|10}(a,a) + \gamma_{11} p_{AB|11}(a, a \oplus 1)\Big) \quad (3.2)$$

where $\gamma_{ij} \in \mathbb{R}$ are some coefficients. However when it comes to analysis of the protocols, we restrict ourselves to the protocols based only on the CHSH tests.

3.3 THE PROGRESS OF DIRNE PROTOCOLS: A BRIEF REVIEW

Before discussing the DIRNE protocols, let us briefly highlight some foundational literature on their development.

The first DIRNE protocol was introduced by Colbeck [1, 2] based on the GHZ test. This work introduced the idea of private randomness generation certified by non-local games. Building on this work, Pironio et al. [3] presented a construction of DIRNE protocol along with its experimental demonstration (this was not a loophole-free experimental demonstration). Further, it was then proven that unbounded randomness can be produced from a finite initial seed of random numbers using many copies of entangled states [37, 48]. Other works have focused on amplification of randomness from an initial source of weak randomness [49].

Meanwhile, improving the security proof of the DIRNE protocols started to attract attention. Pironio and Massar demonstrated that the DIRNE protocol is secure against classical adversaries [50]. This was followed by Vazirani and Vidick, who proved the security of the protocol against an entangled adversary [36]; however, this was without any noise tolerance. Miller and Shi then proved the security in the presence of errors [38]. However, challenges remained for computing reasonable tight bounds on the min-entropy (which quantifies the randomness in a DIRNE protocol – see Chapter 2) for the outputs of the protocol. The difficulty stems from the fact that one needs to account for different collective attacks that a potential adversary may attempt to tamper with the protocol.

Assuming the protocol is carried out in an i.i.d. fashion, the min-entropy can be bound using the single round min-entropy [51] at a finite level of the NPA hierarchy [52, 53]. However, these bounds are generally not very tight. This issue was addressed by the Asymptotic Equipartition Theorem (AEP) [29], which bounds the min-entropy of the protocol in terms of the single round von Neumann entropy. This bound holds for i.i.d. protocols and also in the limit that the number of rounds is large. As von Neumann entropy is larger than the min-entropy, the AEP improves the randomness rates of DIRNE protocols. Therefore, the main challenge was then reduced to getting bounds on the min-entropy of the protocol, keeping two aspects in mind: the i.i.d. assumption needs to be relaxed and the second aspect is to get bounds for finitely many rounds.

Two approaches were developed to address this – the quantum probability estimation framework [13] and the entropy accumulation theorem (EAT) [14, 15]. The EAT bounds the min-entropy in terms of the von Neumann entropy, just like AEP does – however it also provides bounds for finite rounds and for the non-i.i.d. scenario by introducing a penalty term that vanishes as the number of rounds increases. The penalty term in the original EAT was not tight and this EAT bound was further refined in [32] and subsequently in [31]. A generalized version of the EAT was published very recently [30], which extends EAT to cases where the side information of the adversary can be updated after every EAT round.

Although the single-round von Neumann entropy is significantly easier to compute compared to the min-entropy of the entire protocol, the bounds of the single-round von Neumann entropy are hard to compute even for a single round. The challenge arises due to the non-linearity of von Neumann entropy and the lack of assumptions about the states and measurements used. For a vast majority of research up until recently, the randomness rate was determined using known tight bounds on the randomness of just one of the devices. In their work, Pironio et al. [42] calculated bounds on the randomness of a single device in the context of computing key rates for Device Independent Quantum Key Distribution (DIQKD). Since this represents a bound on the randomness of one device, we refer to such a bound as a one-sided rate. As the output of one device is a binary, at most one bit of randomness could be certified using this method. One-sided rates, as opposed to two-sided rates, are useful for DIQKD because the secret key is formed from the outputs of only one party. The other party then performs error correction to agree with the secret key held by the first party. In the literature, this DIQKD bound was re-used to determine a lower bound on the output randomness of the protocol. The advantage was that bounds existed for all values of the CHSH score, allowing for a

non-zero randomness rate as long as there was a violation of the Bell inequality. This result was employed in one of the first loophole-free experimental demonstrations of DIRNE protocols [31], which was conducted recently.

Later, it was shown that in principle, up to 2 bits of randomness could be achieved in a (2-input, 2-output, 2-party) DIRNE protocol using tilted Bell inequalities [44]. This result showed the existence of such strategies, but did not compute the randomness rates for all values of Bell violation. One-sided rates for such tilted Bell inequalities (introduced in the context of DIQKD protocols) for all values of Bell violation were presented by Woodhead et al. [45]. A significant limitation of using such tilted Bell inequalities is that 2 bits can only be certified while being arbitrarily close to the local set (i.e., the CHSH score will be near 2), making high randomness rates less robust against noise.

Therefore, determining bounds on randomness rates for DIRNE protocols (i.e., two-sided bounds) instead of re-using existing one-sided bounds for DIQKD is one the main motivation for this thesis. While working on this thesis, several other important works were also done in this direction, each attempting to address the challenge of finding lower bounds on one-sided and two-sided randomness rates. This led to important advances in computing these bounds [54–56]. More recently, a breakthrough paper [43, 57] was published, presenting a method to compute reliable lower bounds on the von Neumann entropy for the most general non-local games. This technique relies on semi-definite programming, leveraging the NPA hierarchy and methods from non-commutative polynomial optimization theory to derive reliable bounds on von Neumann entropy. Building on this, the recent work by Wooltorton et al. [46] demonstrated that up to two bits of randomness can be extracted even when the observed correlations are far from the local bound (up to a CHSH score of about 2.6). Hence, up to 2 bits of randomness can be extracted from a DIRNE protocol in a more robust manner.

The intriguing question of the quantity of randomness that can be extracted from a single source through multiple measurements was tackled by Curchod et al. Their study [58] revealed that by employing the same initial state and conducting repeated measurements, an unbounded amount of randomness can be extracted. Their method utilized tilted Bell inequalities to certify this randomness. Even more recently, Brown and Colbeck [59] demonstrated that the CHSH inequality can be violated an unbounded number of times, suggesting the potential for extracting unbounded randomness through numerous violations of the CHSH inequality.

Ongoing advancements in theoretical analysis have yielded considerable enhancements in DIRNE protocols. Coupled with the rapid progress observed in experiments [7, 8, 31],

DIRNE protocols are getting increasingly more practical.

3.4 THE SIGNIFICANCE OF VARIOUS ENTROPIC QUANTITIES

In this section, we discuss the significance of the six entropic quantities given above in the context of DIRNE, noting that the one-sided quantities are also useful for DIQKD (Device Independent Quantum Key Distribution). To do so, we first describe the general structure of the raw randomness generation part of a spot-checking and non-spot-checking DIRNE protocol. A more complete description of the protocols is in Chapter 6.

In a protocol without spot-checking (such as the protocol which recycles input randomness), two untrusted devices are used. In every round, their inputs X_i and Y_i are generated according to some distribution p_{XY}. Often two independent random number generators are used for this, so that $p_{XY} = p_X p_Y$. The generated numbers are used as inputs to the devices, which return two outputs A_i and B_i respectively. This is repeated for n rounds generating the raw randomness $\mathbf{A}$, $\mathbf{B}$, where the bold font denotes the concatenation of all the outputs.

In a spot-checking protocol, there is an added step. In this step, each round is declared either a test round ($T_i = 1$) or a generation round ($T_i = 0$). Test rounds occur with a typically small probability γ. In test rounds, X_i and Y_i are generated according to some distributions p_{XY}. In generation rounds, X_i and Y_i are set according to some other distributions – in this work we use the deterministic distribution $X_i = Y_i = 0$. These are used as inputs to the devices, which return two outputs A_i and B_i respectively. The rationale behind using a spot-checking protocol is that randomness is required to perform a Bell test and it is desirable to be able to run the protocol with a smaller requirement on the amount of input randomness required. Choosing whether to test or not requires roughly $H_{\text{bin}}(\gamma)$ bits of randomness per round[4], so choosing γ small enough leads to an overall saving. Furthermore, protocols often discard the input randomness, in which case for many Bell tests spot-checking is necessary in order to achieve expansion. In the CHSH game, for instance, if p_{XY} is chosen uniformly, each test round requires 2 bits of randomness, but the amount of two-sided randomness output by the quantum strategy with the highest possible winning probability is only $1 + H_{\text{bin}}(\frac{1}{2}(1 + \frac{1}{\sqrt{2}})) \approx 1.60$ bits. However, as we discuss later, the input randomness need not be discarded.

[4]Here $H_{\text{bin}} : [0,1] \mapsto [0,1]$ denotes the binary entropy defined as $H_{\text{bin}}(x) = -x\log_2(x) - (1-x)\log_2(1-x)$. Here $\log_2$ is the logarithm with base 2. Further, $0\log_2 0$ is taken to be 0.

In the case of small γ, almost every round is a generation round hence an eavesdropper wishes to guess the outputs for the inputs $X = 0$ and $Y = 0$. The entropy $H(AB|X = 0, Y = 0, E)$ is thus the relevant quantity for spot-checking DIRNE protocols. The one-sided quantity $H(A|X = 0, Y = 0, E)$ has often been used instead because of the existing analytic bound for this [42, 45], but, because this ignores one of the outputs, it is wasteful as an estimate of the generated randomness. For DIQKD protocols, on the other hand, the one-sided entropy is the relevant quantity. This is because in order to make a key, the random strings held by Alice and Bob, $\mathbf{A}$ and $\mathbf{B}$ respectively should match. Thus, only one of the strings $\mathbf{A}$ can be used as a key, and the other string $\mathbf{B}$ should be corrected to match the with the string $\mathbf{A}$. We also remark that these quantities can be useful bounds for protocols without spot-checking if the distribution p_{XY} is heavily biased towards $X = Y = 0$.

The quantities $H(AB|XYE)$ and $H(A|XYE)$ are useful for protocols without spot checking. One might imagine, for example, using a source of public randomness, such as a randomness beacon to choose the inputs to the protocol, in which case X and Y become known to the adversary (but are not known before the devices are prepared). In this case, rather than being interested in randomness expansion, the task is to turn public randomness into private randomness in a Device Independent way. One can also use $H(AB|XYE)$ and $H(A|XYE)$ in protocols when the input randomness is recycled. In this case we are really interested in $H(ABXY|E)$, but, because X and Y are chosen independently of E, this can be expanded as $H(XY) + H(AB|XYE)$. Hence $H(AB|XYE)$ is the relevant quantity in this case as well. The one-sided quantity $H(A|XYE)$ could also be used for DIQKD without spot-checking.

In addition we consider the quantities $H(AB|E)$ and $H(A|E)$. The second of these could be useful for QKD protocols in which the key generation rounds do not have a fixed input and where Alice and Bob do not publicly reveal their measurement choices during the protocol. For instance, the sharing of these choices could be encrypted using an initial key, analogously to a suggested defence against memory attacks [60][5]. $H(AB|E)$ would be a useful quantity for randomness generation in a protocol without spot-checking and in which X and Y are kept private after running the protocol and not used in the overall output. When such protocols are based on the CHSH inequality, they cannot allow expansion. These quantities can also be thought of as quantifying the

[5] Note that such protocols would only be useful if more key is generated than is required, so the protocol we are thinking of here is really quantum key expansion. Furthermore, the results presented in Figure 4.1 show that the use of $H(A|E)$ only gives a minor advantage over $H(A|XYE)$.

fundamental amount of randomness obtainable from a given Bell violation. Although we have computed the graphs for $H(AB|E)$ and $H(A|E)$, existing versions of the EAT cannot be directly applied to them — see Appendix A.4.

— 4 —

Upper bounds on the entropic quantities

In the previous chapter, we noted the necessity for computing lower bounds on entropic quantities $H(.|.E)$ for any strategy that achieves a CHSH score ω when carried out in an i.i.d. manner. This computation is crucial for determining the rates of CHSH-based protocols. In this chapter, we proceed with a rigorous definition of such a strategy, formulating it as an optimization problem for the entropic quantities. Techniques such as Jordan's lemma will be employed to simplify these optimization problems, enabling us to calculate the asymptotic rates of the protocols (i.e., the randomness per round). Interestingly, we observe that our techniques have a broader applicability and can be extended to accommodate a wider class of Bell inequalities, of which the CHSH inequality is a special case. We refer to these inequalities as CHSH-type inequalities and will define them in the subsequent section. Nonetheless, for our analysis, we primarily focus on the CHSH inequality due to its widespread use in literature concerning randomness expansion protocols.

Parts of Chapter 4, including Sections 4.2, 4.3, 4.4, 4.5, and 4.9, build upon techniques presented in the work by Pironio et al. [42]. Their work was dedicated to finding tight bounds on the one-sided rate $H(A|X=0,Y=0,E)$ as a function of the CHSH score ω and was derived for a DIQKD protocol. In this thesis, we extend their approach and present it within a more mathematically rigorous framework to make the approach taken compatible with the framework of the entropy accumulation theorem. Moreover, we extend these results to compute the bounds for all one and two-sided entropies $H(.|.E)$ as a function of the CHSH score ω (refer to Section 3.4 for the appropriate definitions). We also show that these techniques can further be extended to find tight bounds for all the 6 one- and two-sided rates as a function of the generalized CHSH score as defined in equation 3.2.

4.1 RATES FOR (GENERALIZED) CHSH-BASED PROTOCOLS

We calculate various one-sided and two-sided rates for protocols based on the (generalized) CHSH game, which involves trying to violate the CHSH Bell inequality [61]. Recall that in this game, each party uses a binary input and receives a binary output and the game is won if $A \oplus B = XY$. We define the generalized CHSH score by

$$\begin{aligned}\omega \;:=\; & \gamma_{00}\sum_a p_{AB|00}(a,a) + \gamma_{01}\sum_a p_{AB|01}(a,a) + \gamma_{10}\sum_a p_{AB|10}(a,a) \\ & + \gamma_{11}\sum_a p_{AB|11}(a,a\oplus 1), \end{aligned} \tag{4.1}$$

which is the probability of winning the generalized CHSH game when the inputs are chosen at random[1]. Note here that γ_{ij} are some real coefficients. A special feature of this class of scores is that they remain invariant under the re-labelling of the output variables $(a,b) \to (a\oplus 1, b\oplus 1)$. As we shall see later in this chapter, having such a relabelling symmetry implies that there are always two different strategies (related by a deterministic local operation) that yield the same score. This property will be leveraged to simplify the optimization problem for computing the randomness rates as a function of the score ω.

The CHSH score is the special case when all the coefficients $\gamma_{ij} = \frac{1}{4}$. Classical strategies the in case of the protocols based on the CHSH inequality can win this game with probability at most $3/4$, while quantum strategies can get as high as $\frac{1}{2}\left(1+\frac{1}{\sqrt{2}}\right) \approx 0.85$. For a fixed generalized CHSH score, ω, we wish to compute the minimum von Neumann entropy over all strategies achieving that score. In this context, a strategy comprises three Hilbert spaces $\mathcal{H}_{A'}$, $\mathcal{H}_{B'}$ and $\mathcal{H}_E$, POVMs $\{M_{a|x}\}_a$ on $\mathcal{H}_{A'}$ for both $x=0$ and $x=1$, POVMs $\{N_{b|y}\}_b$ on $\mathcal{H}_{B'}$ for both $y=0$ and $y=1$, and a state $\rho_{A'B'E}$ on $\mathcal{H}_{A'}\otimes\mathcal{H}_{B'}\otimes\mathcal{H}_E$. Given a strategy and a distribution p_{XY} there is an associated channel $\mathcal{N}$ that acts on $A'B'$ and takes the state to the post-measurement state, i.e.,

$$\begin{aligned}\tau_{ABXYE} &= (\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E}) \\ &= \sum_{abxy} p_{XY}(x,y)\,|abxy\rangle\langle abxy|_{A,B,X,Y} \otimes \mathrm{tr}_{A'B'}\left((M_{a|x}\otimes N_{b|y}\otimes \mathbb{1}_E)\rho_{A'B'E}\right),\end{aligned}$$

where $\mathcal{I}_E$ is the identity channel on E. The entropic quantities we consider all pertain to this state[2]. Note also that the generalized CHSH score that is achieved using the

[1] Note that even if nonuniform distributions of inputs are used when running protocols, in this work the CHSH score is always defined as (4.1).

[2] In the cases where we condition on $X=0$ and $Y=0$, we can project this state onto $|0\rangle\langle 0|_X \otimes |0\rangle\langle 0|_Y$ and renormalize — see section 4.5 for more detail.

strategy (if it is performed in an i.i.d. fashion) is a function of τ, which we denote by $S(\tau)$ – in particular $S(\tau)$ is defined using Equation (4.1) and substituting $p_{AB|xy}(a,b)$ by the expression $\mathrm{tr}\left((M_{a|x}\otimes N_{b|y}\otimes \mathbb{1}_E)\rho_{A'B'E}\right)$.

For each of the six entropic quantities previously discussed, we consider the infimum over all strategies that achieve a given score. We use this to define a set of curves corresponding to each of the 6 entropic quantities. We write $F_{AB|XYE}(\omega, p_{XY}) = \inf H(AB|XYE)_\tau$, where the infimum is over all strategies for which $S(\tau)=\omega$. In the same way we define $F_{AB|E}(\omega, p_{XY})$, $F_{A|XYE}(\omega, p_{XY})$ and $F_{A|E}(\omega, p_{XY})$ by replacing the objective function with the corresponding entropy. We also define $F_{AB|00E}(\omega)$ and $F_{A|00E}(\omega)$ analogously, noting that these are independent of p_{XY}. Furthermore, if we write $F_{AB|XYE}(\omega)$ etc. (i.e., leaving out the p_{XY}), we refer to the case where p_{XY} is uniform over X and Y. For a more precise writing of these optimizations, see Equation (4.8).

We also consider a related set of functions $G_{AB|XYE}(\omega, p_{XY})$, $G_{AB|E}(\omega, p_{XY})$ etc. that are defined analogously, but while optimizing over a smaller set of allowed strategies. More precisely, the G functions are defined by restricting $\mathcal{H}_{A'}$ and $\mathcal{H}_{B'}$ to be two dimensional and $\mathcal{H}_E$ to have dimension 4, taking $\rho_{A'B'E}$ to be pure with $\rho_{A'B'}$ diagonal in the Bell basis, and taking the POVMs to be projective measurements onto states of the form $\cos(\alpha)\,|0\rangle + \sin(\alpha)\,|1\rangle$ (see Equation (4.14)).

Note that ideally, the family of functions $F_{.|.E}(\omega, p_{XY})$ and $G_{.|,E}(\omega, p_{XY})$ should explicitly indicate the coefficients γ_{ij} used to define the score. However, for the sake of brevity and to avoid unnecessary complexity in notation, we will omit this explicit dependence. In the later parts of this chapter and the entirety of the next chapter, we will use ω to refer exclusively to the CHSH score.

As we shall see later in this chapter, it turns out that $F_{AB|00E}(\omega) = G_{AB|00E}(\omega)$, $F_{A|00E}(\omega) = G_{A|00E}(\omega)$, and that in each of the other four cases F and G are related by

$$F = \mathrm{convenv}(G). \tag{4.2}$$

Here $\mathrm{convenv}(G)$ represents the convex envelope (or the convex lower bound) of G. Roughly speaking, the convex envelope of a function g is the smallest convex function f that is not greater than g. The rigorous definition of this function in its most general form can be found in Section 8.10. The underlying reason why the above equation (eqn. 4.2) holds is due to the Jordan's lemma. In our context, Jordan's lemma [41] implies that in the case of Bell inequalities with two inputs and two outputs, any strategy is equivalent to a convex combination of strategies in which A' and B' are qubit systems. This means that if we solve the qubit case, the general case follows by taking the convex

lower bound. Such an argument was made in [42] and we now proceed to show this in the next few sections.

A note on notation: in this work we measure entropies in bits, taking $\log$ to represent the logarithm base 2, and $\ln$ for the natural logarithm where needed.

4.2 SIMPLIFYING THE STRATEGY

Given a Hilbert space $\mathcal{H}$, we define $\mathcal{P}(\mathcal{H})$ to be the set of positive semi-definite operators on $\mathcal{H}$, and $\mathcal{S}(\mathcal{H})$ to be the set of density operators, i.e., elements of $\mathcal{P}$ with trace 1. The pure states on $\mathcal{H}$ (elements of $\mathcal{S}(\mathcal{H})$ with rank 1) will be denoted $\mathcal{S}_P(\mathcal{H})$. A POVM on $\mathcal{H}$ is a set of positive operators $\{E_i\}_i$ with $E_i \in \mathcal{P}(\mathcal{H})$ for all i and $\sum_i E_i = \mathbb{1}_{\mathcal{H}}$, where $\mathbb{1}_{\mathcal{H}}$ is the identity operator on $\mathcal{H}$. A projective measurement on $\mathcal{H}$ is a POVM on $\mathcal{H}$ where $E_i^2 = E_i$ for all i. We define the Bell states

$$|\Phi_0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{4.3}$$

$$|\Phi_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{4.4}$$

$$|\Phi_2\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{4.5}$$

$$|\Phi_3\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)\ , \tag{4.6}$$

and use $\sigma_1 = |1\rangle\langle 0| + |0\rangle\langle 1|$, $\sigma_2 = \mathrm{i}\,|1\rangle\langle 0| - \mathrm{i}\,|0\rangle\langle 1|$ and $\sigma_3 = |0\rangle\langle 0| - |1\rangle\langle 1|$ as the three Pauli operators.

In this section we make a series of simplifications of the form of the optimization. The argument given broadly follows the logic of [42] (see also [45] for an alternative).

Definition 13. A *single-round* $2-2-2$ *measurement strategy* is a tuple $(\mathcal{H}_{A'}, \mathcal{H}_{B'}, \{M_{a|x}\}_{x,a}, \{N_{b|y}\}_{y,b})$, where $\mathcal{H}_{A'}$ and $\mathcal{H}_{B'}$ are Hilbert spaces, and $\{M_{a|x}\}_a$ is a POVM on $\mathcal{H}_{A'}$ for each $x \in \{0,1\}$ and likewise $\{N_{b|y}\}_b$ is a POVM on $\mathcal{H}_{B'}$ for each $y \in \{0,1\}$. In the case that all the POVMs are projective we will call this a *single-round* $2-2-2$ *projective measurement strategy*.

Note that here $2-2-2$ stands for 2 possible values of inputs, 2 possible values for the outputs and 2 parties.

Definition 14. A *single-round* $2-2-2$ *strategy* is a single-round $2-2-2$ measurement strategy together with a state $\rho_{A'B'E} \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E$, where $\mathcal{H}_E$ is an arbitrary Hilbert space.

Note that in a device-independent scenario, such a strategy can be chosen by the adversary.

Definition 15. Given a single-round $2-2-2$ measurement strategy and a distribution p_{XY} over the settings X and Y, the *associated* $2-2-2$ *channel* is defined by

$$\mathcal{N} : \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_X \otimes \mathcal{H}_Y) :$$
$$\sigma \mapsto \sum_{abxy} p_{XY}(x,y)\, |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes |x\rangle\langle x| \otimes |y\rangle\langle y| \operatorname{tr}\left((M_{a|x} \otimes N_{b|y})\sigma\right) ,$$

where $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_X$ and $\mathcal{H}_Y$ are two dimensional Hilbert spaces. The union of the sets of associated $2-2-2$ channels for all single-round $2-2-2$ measurement strategies for some fixed input distribution p_{XY} is denoted $\mathcal{C}(p_{XY})$. The union of the sets of associated $2-2-2$ channels for all single-round $2-2-2$ projective measurement strategies is denoted $\mathcal{C}_\Pi(p_{XY})$.

The output of the associated $2-2-2$ channel is classical, and $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_X \otimes \mathcal{H}_Y$ stores the outcomes and the chosen measurements. We will usually apply this channel to the AB part of a tripartite system, giving

$$(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) = \sum_{abxy} p_{XY}(x,y) p_{AB|xy}(a,b)\, |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \tau_E^{a,b,x,y} , \tag{4.7}$$

where $\tau_E^{a,b,x,y} \in \mathcal{S}(\mathcal{H}_E)$ for each a, b, x, y (it is the normalization of $\operatorname{tr}_{A'B'}\left((M_{a|x} \otimes N_{b|y} \otimes \mathbb{1}_E)\rho_{A'B'E}\right)$).
Note that the generalized CHSH score, which we denote $S((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}))$ does not depend on the distribution p_{XY} of input settings.
We will be interested in optimization problems of the form

$$F(\omega, p_{XY}) = \inf_{\mathcal{R}} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})), \text{ where} \tag{4.8}$$
$$\mathcal{R} = \{(\mathcal{N}, \rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}(p_{XY}),\ S((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \omega\},$$

where $\mathcal{H}_E$ is an arbitrary Hilbert space and the spaces $\mathcal{H}_{A'}$ and $\mathcal{H}_{B'}$ are those from the chosen element of $\mathcal{C}(p_{XY})$, i.e., the set $\mathcal{R}(\omega)$ runs over all possible dimensions of these spaces, and ω is some fixed real number. Here $\bar{H}$ can be any one of the following

entropic quantities defined on the state $(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$: $H(AB|X=0,Y=0,E)$, $H(AB|XYE)$, $H(AB|E)$, $H(A|X=0,Y=0,E)$, $H(A|XYE)$ or $H(A|E)$. We consider the family of optimizations in this work and many of the arguments that follow to be independent of this choice.

4.2.1 Reduction to projective measurements

In this section, we conclude that there is no loss in generality in assuming that the devices perform projective measurements. More precisely, we prove the following lemma.

Lemma 1. *The sets*

$$\begin{aligned}\mathcal{T}_1 &:= \{(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}(p_{XY}), \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)\} \text{ and}\\ \mathcal{T}_2 &:= \{(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_\Pi(p_{XY}), \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)\}\end{aligned}$$

are identical.

This is a corollary of Naimark's theorem, which we state in the following way.

Theorem 2 (Naimark's theorem)**.** *Let $\{E_i\}_i$ be a POVM on $\mathcal{H}$. There exists a Hilbert space $\mathcal{H}'$ and a projective measurement $\{\Pi_i\}_i$ on $\mathcal{H} \otimes \mathcal{H}'$ such that for any $\rho \in \mathcal{S}(\mathcal{H})$*

$$\sum_i |i\rangle\langle i| \operatorname{tr}(\rho E_i) = \sum_i |i\rangle\langle i| \operatorname{tr}(\Pi_i(\rho \otimes |0\rangle\langle 0|)) .$$

Proof. We can directly construct this measurement as follows. Consider the isometry $V : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}'$ given by $V = \sum_i \sqrt{E_i} \otimes |i\rangle$, and let U be the extension of V to a unitary with the property that $U(|\psi\rangle \otimes |0\rangle) = \sum_i \sqrt{E_i}\,|\psi\rangle \otimes |i\rangle$ for any $|\psi\rangle \in \mathcal{H}$. This construction ensures that the channels

$$\begin{aligned}\mathcal{E} : \rho &\mapsto \sum_i |i\rangle\langle i| \operatorname{tr}(E_i \rho) \text{ and}\\ \mathcal{E}' : \rho &\mapsto \sum_i |i\rangle\langle i| \operatorname{tr}\left((\mathbb{1} \otimes |i\rangle\langle i|) U (\rho \otimes |0\rangle\langle 0|) U^\dagger\right)\end{aligned}$$

are identical. The second of these can be rewritten

$$\rho \mapsto \sum_i |i\rangle\langle i| \operatorname{tr}\left(\Pi_i(\rho \otimes |0\rangle\langle 0|)\right) ,$$

where we take $\Pi_i = U^\dagger(\mathbb{1} \otimes |i\rangle\langle i|)U$, as required. □

Proof of Lemma 1. By definition $\mathcal{T}_2 \subseteq \mathcal{T}_1$. For the other direction, consider a state $\rho_{A'B'E}$ and POVMs $\{M_{a|x}\}_{a,x}$ and $\{N_{b|y}\}_{b,y}$ forming a single-round CHSH strategy in $\mathcal{T}_1$. We use the construction in the proof of Theorem 2 to generate the projectors $\Pi^A_{a|x}$ and $\Pi^B_{b|y}$ as Naimark extensions of the POVMs. Instead of creating the state $\rho_{A'B'E}$, the state $\rho_{A'B'E} \otimes |0\rangle\langle 0|_{A''} \otimes |0\rangle\langle 0|_{B''}$ is created, where the projectors $\Pi^A_{a|x}$ act on $A'A''$ and $\Pi^B_{b|y}$ act on $B'B''$. Since the latter is a strategy in $\mathcal{T}_2$ leading to the same post-measurement state (4.7), we have $\mathcal{T}_1 \subseteq \mathcal{T}_2$, which completes the proof. □

4.2.2 Reduction to convex combinations of qubit strategies

This is a consequence of Jordan's lemma [41] and is a special feature that applies only because the Bell inequality has two inputs and two outputs for each party.

Lemma 2 (Jordan's lemma)**.** *Let A_1 and A_2 be two Hermitian operators on $\mathcal{H}$ with eigenvalues ± 1, then we can decompose $\mathcal{H} = \bigoplus_\alpha \mathcal{H}_\alpha$ such that A_1 and A_2 preserve the subspaces $\mathcal{H}_\alpha$, and where each $\mathcal{H}_\alpha$ has dimension at most 2.*

Corollary 1. *Let Π_1 and Π_2 be two projections on $\mathcal{H}$. We can decompose $\mathcal{H} = \bigoplus_\alpha \mathcal{H}_\alpha$ such that Π_1, $\mathbb{1} - \Pi_1$, Π_2 and $\mathbb{1} - \Pi_2$ preserve the subspaces $\mathcal{H}_\alpha$, and where each $\mathcal{H}_\alpha$ has dimension at most 2.*

Proof. Apply Jordan's lemma to the Hermitian operators $A_1 = 2\Pi_1 - \mathbb{1}$ and $A_2 = 2\Pi_2 - \mathbb{1}$ with eigenvalues ± 1, and consider $|\psi\rangle \in \mathcal{H}_\alpha$ for some α. By construction $A_1 |\psi\rangle \in \mathcal{H}_\alpha$ from which it follows that $\Pi_1 |\psi\rangle \in \mathcal{H}_\alpha$, and hence also $(\mathbb{1} - \Pi_1) |\psi\rangle \in \mathcal{H}_\alpha$. Thus, Π_1 and $\mathbb{1} - \Pi_1$ preserve the subspace; likewise Π_2 and $\mathbb{1} - \Pi_2$. □

This implies the following

Lemma 3. *Let $\mathcal{C}_{2\times 2}(p_{XY})$ be the set of CHSH channels associated with the single-round CHSH projective measurement strategies where each of the four projectors $M_{a|x}$ is block diagonal with 2×2 blocks, and each of the four projectors $N_{b|y}$ is block diagonal with 2×2 blocks. The sets*

$$\mathcal{T}_2 := \{(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_\Pi(p_{XY}), \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)\} \text{ and}$$
$$\mathcal{T}_3 := \{(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_{2\times 2}(p_{XY}), \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)\}$$

are identical.

Proof. This follows by applying Corollary 1 to the projectors $M_{0|0}$ and $M_{0|1}$ to get the blocks on $\mathcal{H}_{A'}$ and to the projectors $N_{0|0}$ and $N_{0|1}$ to get the blocks on $\mathcal{H}_{B'}$. Although some of the blocks may be 1×1, we can collect these together and treat them as a 2×2 block, or add an extra dimension to the space (on which the state has no support) to achieve all 2×2 blocks. □

We can also make the state only have support on the 2×2 blocks.

Lemma 4. *The sets*

$$\begin{aligned} \mathcal{T}_3 &:= \{(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_{2\times 2}(p_{XY}), \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)\} \text{ and} \\ \mathcal{T}_4 &:= \{(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_{2\times 2}(p_{XY}), \rho_{A'B'E} \in \mathcal{S}_{2\times 2}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)\} \end{aligned}$$

are identical. Here $\mathcal{S}_{2\times 2}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)$ is the subset of $\mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)$ such that $\rho_{A'B'E} \in \mathcal{S}_{2\times 2}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)$ implies

$$\rho_{A'B'E} = \sum_{\alpha,\beta} (\Pi^\alpha_{A'} \otimes \Pi^\beta_{B'} \otimes \mathbb{1}_E) \rho_{A'B'E} (\Pi^\alpha_{A'} \otimes \Pi^\beta_{B'} \otimes \mathbb{1}_E) \,,$$

where $\{\Pi^\alpha\}_\alpha$ are projectors onto the 2×2 diagonal blocks.

Proof. Consider a state $\rho_{A'B'E}$ and sets of projectors $\{M_{a|x}\}_{a,x}$ and $\{N_{b|y}\}_{b,y}$ from the set $\mathcal{T}_3$. For brevity, write $\Pi^{\alpha,\beta} = \Pi^\alpha_{A'} \otimes \Pi^\beta_{B'}$. Then, since

$$M_{a|x} \otimes N_{b|y} = \sum_{\alpha,\beta} (\Pi^\alpha_{A'} \otimes \Pi^\beta_{B'})(M_{a|x} \otimes N_{b|y})(\Pi^\alpha_{A'} \otimes \Pi^\beta_{B'}) \,,$$

we have

$$\begin{aligned} \mathrm{tr}\Big((M_{a|x} \otimes N_{b|y} \otimes \mathbb{1})\rho_{A'B'E}\Big) &= \mathrm{tr}\left(\sum_{\alpha,\beta} (\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)(M_{a|x} \otimes N_{b|y} \otimes \mathbb{1})(\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)\rho_{A'B'E}\right) \\ &= \mathrm{tr}\left(\sum_{\alpha,\beta} (M_{a|x} \otimes N_{b|y} \otimes \mathbb{1})(\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)\rho_{A'B'E}(\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)\right) \\ &= \mathrm{tr}_{A'B'}((M_{a|x} \otimes N_{b|y} \otimes \mathbb{1})\rho'_{A'B'E}) \,, \end{aligned}$$

where $\rho'_{A'B'E} = \sum_{\alpha,\beta} (\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)\rho_{A'B'E}(\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)$ and tr is over the registers A' and B'. Thus, if we replace $\rho_{A'B'E}$ by $\rho'_{A'B'E}$ we obtain the same post-measurement state (4.7). Hence $\mathcal{T}_3 \subseteq \mathcal{T}_4$, and, since the other inclusion is trivial, $\mathcal{T}_3 = \mathcal{T}_4$. □

Lemma 5. *Let $\mathcal{N} \in \mathcal{C}_{2\times 2}(p_{XY})$ and $\rho_{A'B'E} \in \mathcal{S}_{2\times 2}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)$. The state $(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$ can be formed as a convex combination of states $(\mathcal{N}_\lambda \otimes \mathcal{I}_E)(\rho^\lambda_{A''B''E})$, where for each λ, the channel $\mathcal{N}_\lambda$ is that associated with a single-round measurement strategy with two 2-dimensional Hilbert spaces and distribution p_{XY}.*

Proof. Since $\rho_{A'B'E} \in \mathcal{S}_{2\times 2}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)$ the 2×2 block structure means we can write $\rho_{A'B'E} = \sum_{\alpha,\beta} p_{\alpha,\beta} \rho^{\alpha,\beta}_{A'B'E}$, where $p_{\alpha,\beta}\rho^{\alpha,\beta}_{A'B'E} = (\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)\rho_{A'B'E}(\Pi^{\alpha,\beta} \otimes \mathbb{1}_E)$ and $\mathrm{tr}\left(\rho^{\alpha,\beta}_{A'B'E}\right) = 1$ for all α and β. Likewise, taking $M^{\alpha}_{a|x} = \Pi^{\alpha}_{A'} M_{a|x} \Pi^{\alpha}_{A'}$ and $N^{\beta}_{b|y} = \Pi^{\beta}_{B'} N_{b|y} \Pi^{\beta}_{B'}$ we can write $M_{a|x} = \sum_\alpha M^{\alpha}_{a|x}$ and $N_{b|y} = \sum_\beta N^{\beta}_{b|y}$. In terms of these we have

$$\mathrm{tr}_{A'B'}((M_{a|x} \otimes N_{b|y} \otimes \mathbb{1})\rho_{A'B'E}) \;=\; \sum_{\alpha,\beta} p_{\alpha,\beta}\, \mathrm{tr}_{A'B'}((M^{\alpha}_{a|x} \otimes N^{\beta}_{b|y} \otimes \mathbb{1})\rho^{\alpha,\beta}_{A'B'E}).$$

We can then associate a value of λ with each pair (α, β), replace each $\rho^{\alpha,\beta}_{A'B'E}$ by a state on $A''B''E$ in which A'' and B'' are two-dimensional (the support of $\rho^{\alpha,\beta}_{A'B'}$ has dimension at most 4), and likewise replace the projectors by qubit projectors. In terms of these we have

$$(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) = \sum_\lambda p_\lambda (\mathcal{N}_\lambda \otimes \mathcal{I}_E)(\rho^{\lambda}_{A''B''E})\,. \qquad \square$$

In other words, any post-measurement state (4.7) that can be generated in the general case, can also be generated if Eve sends a convex combination of two qubit states, and where the measurements used by the separated devices depend on the state sent. Eve could realise such a strategy in practice by using pre-shared randomness. We can proceed to consider strategies in which qubits are shared between the two devices, and then consider the mixture of such strategies after doing so.

4.3 QUBIT STRATEGIES

In this section we consider the single-round $2-2-2$ measurement strategies in which $\mathcal{H}_{A'}$ and $\mathcal{H}_{B'}$ are two-dimensional and the measurements are rank-1 projectors. Given a distribution p_{XY} we use $\mathcal{C}_{\Pi_1,2}(p_{XY})$ to denote the set of associated $2-2-2$ channels. We restrict to rank-1 projectors because if one of the projectors is equal to the identity it is not possible to achieve a non-classical generalized CHSH score, and the non-classical scores are the ones of interest.

Lemma 6. *Consider a single-round $2-2-2$ measurement strategy for which one of the POVM elements is identity and let $\mathcal{N}$ be the associated $2-2-2$ channel. For any state $\rho_{A'B'}$ on which $\mathcal{N}$ can act we have $S(\mathcal{N}(\rho_{A'B'})) \leq l$, where l is the local bound for the generalized CHSH inequality.*

Proof. Suppose the identity element corresponds to $M_{0|0}$ (the other cases follow symmetrically). Any conditional distribution $p_{AB|XY}$ that obeys the non-signalling conditions takes the form

		$X=0$		$X=1$	
		$A=0$	$A=1$	$A=0$	$A=1$
$Y=0$	$B=0$	μ	0	ν	$\mu-\nu$
	$B=1$	$1-\mu$	0	ζ	$1-\mu-\zeta$
$Y=1$	$B=0$	γ	0	ξ	$\gamma-\xi$
	$B=1$	$1-\gamma$	0	$\nu+\zeta-\xi$	$1+\xi-\gamma-\nu-\zeta$

where μ, ν, ζ and γ are any arbitrary non-negative numbers that do not exceed 1. In order to prove the lemma, it suffices to show that this distribution can be achieved using a local strategy. For this, we show that the CHSH score for this strategy $\leq \frac{3}{4}$. The associated CHSH score is

$$\frac{1}{4}\left(\mu+\nu+(1-\mu-\zeta)+\gamma+(\gamma-\xi)+(\nu+\zeta-\xi)\right) \;=\; \frac{1}{4}\left(1+2\nu+2\gamma-2\xi\right).$$

Since every element of the distribution must be between 0 and 1 we have $1+\xi-\gamma-\nu-\zeta \geq 0$, and hence $1+2\nu+2\gamma-2\xi \leq 3-2\zeta \leq 3$, from which the claim follows. □

We will then consider an optimization of the form (4.8), but restricting to $\mathcal{C}_{\Pi_1,2}$, i.e.

$$h(\omega) = \inf_{\mathcal{R}(\omega)} \bar{H}((\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E})), \text{ where} \tag{4.9}$$
$$\mathcal{R}(\omega) = \{(\mathcal{N}, \rho_{A'B'E}) : \mathcal{N}\in\mathcal{C}_{\Pi_1,2}(p_{XY}),\, S((\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E})) = \omega\}.$$

The next step is to show that without loss of generality we can reduce to states that are invariant under application of $\sigma_2\otimes\sigma_2$ on $A'B'$.

Lemma 7. *Let p_{XY} be a distribution, $\mathcal{N}\in\mathcal{C}_{\Pi_1,2}(p_{XY})$ and $\rho_{A'B'E}\in\mathcal{S}(\mathcal{H}_{A'}\otimes\mathcal{H}_{B'}\otimes\mathcal{H}_E)$ be such that $S((\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E})) = \omega$. There exists a state $\tilde{\rho}_{A'B'EE'}\in\mathcal{S}(\mathcal{H}_{A'}\otimes\mathcal{H}_{B'}\otimes\mathcal{H}_E\otimes\mathcal{H}_{E'})$ such that $S((\mathcal{N}\otimes\mathcal{I}_{EE'})(\tilde{\rho}_{A'B'EE'})) = \omega$, $\tilde{\rho}_{A'B'EE'} = (\sigma_2\otimes\sigma_2\otimes\mathbb{1}_{EE'})\tilde{\rho}_{A'B'EE'}(\sigma_2\otimes\sigma_2\otimes\mathbb{1}_{EE'})$ and $\bar{H}((\mathcal{N}\otimes\mathcal{I}_{EE'})(\tilde{\rho}_{A'B'EE'})) = \bar{H}((\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E}))$ for all six of the entropic functions given earlier.*

Note that this implies that $p_{A|X}$ and $p_{B|Y}$ can be taken to be uniform.
This is a consequence of the following lemmas.

Lemma 8. *Let $\{\Pi_{0|0},\Pi_{1|0}\}$ and $\{\Pi_{0|1},\Pi_{1|1}\}$ be two rank-one projective measurements on a two dimensional Hilbert space $\mathcal{H}$. There exists a basis $\{|e_i\rangle\}_{i=1}^2$ such that $\langle e_l|\,\Pi_{i|j}\,|e_k\rangle\in\mathbb{R}$ for all i,j,k,l.*

Proof. Without loss of generality, we can take $\Pi_{0|0} = |0\rangle\langle 0|$ and $\Pi_{1|0} = |1\rangle\langle 1|$, and then write $\Pi_{0|1} = \left|\alpha_{0|1}\right\rangle\left\langle\alpha_{0|1}\right|$ and $\Pi_{1|1} = \left|\alpha_{1|1}\right\rangle\left\langle\alpha_{1|1}\right|$, where $\left|\alpha_{0|1}\right\rangle = \cos(\lambda)\,|0\rangle + \mathrm{e}^{\mathrm{i}\chi}\sin(\lambda)\,|1\rangle$ and $\left|\alpha_{1|1}\right\rangle = \sin(\lambda)\,|0\rangle - \mathrm{e}^{\mathrm{i}\chi}\cos(\lambda)\,|1\rangle$. Then, we can re-define $|1\rangle \to \mathrm{e}^{\mathrm{i}\chi}\,|1\rangle$ so that $\left|\alpha_{0|1}\right\rangle = \cos(\lambda)\,|0\rangle + \sin(\lambda)\,|1\rangle$ and $\left|\alpha_{1|1}\right\rangle = \sin(\lambda)\,|0\rangle - \cos(\lambda)\,|1\rangle$, with $\lambda \in \mathbb{R}$. □

Lemma 9. *Let $\{\Pi_{0|0}, \Pi_{1|0}\}$ and $\{\Pi_{0|1}, \Pi_{1|1}\}$ be two rank-one projective measurements on a two dimensional Hilbert space $\mathcal{H}$, then, there exists a unitary transformation U such that $U\Pi_{j|i}U^\dagger = \Pi_{j\oplus 1|i}$ for all i, j.*

Proof. Let $\Pi_{j|i} = \left|\alpha_{0|1}\right\rangle\left\langle\alpha_{0|1}\right|$ for all i, j. From Lemma 8, we can change basis such that $\left|\alpha_{0|0}\right\rangle = |0\rangle$, $\left|\alpha_{1|0}\right\rangle = |1\rangle$, $\left|\alpha_{0|1}\right\rangle = \cos(\lambda)\,|0\rangle + \sin(\lambda)\,|1\rangle$ and $\left|\alpha_{1|1}\right\rangle = \sin(\lambda)\,|0\rangle - \cos(\lambda)\,|1\rangle$ for some $\lambda \in \mathbb{R}$. Any unitary of the form $U = \mathrm{e}^{\mathrm{i}\phi}(|0\rangle\langle 1| - |1\rangle\langle 0|)$, with $\phi \in \mathbb{R}$ then satisfies the desired relations. □

The following lemma is well-known (it follows straightforwardly from e.g., [62, Section 11.3.5])

Lemma 10. *For $\rho_{CZEE'} = \sum_i p_i \rho^i_{CZE} \otimes |i\rangle\langle i|_{E'}$ we have $H(C|ZEE')_\rho = \sum_i p_i H(C|ZE)_{\rho^i}$.*

We now prove Lemma 7.

Proof of Lemma 7. Let U_A and U_B be the unitaries formed by applying Lemma 9 to respective measurements of each device and using the choice of basis specified in the proof of Lemma 9 we can take $U_A = \sigma_2$ and $U_B = \sigma_2$. Then define $\rho'_{A'B'E} = (\sigma_2 \otimes \sigma_2 \otimes \mathbb{1}_E)\rho_{A'B'E}(\sigma_2 \otimes \sigma_2 \otimes \mathbb{1}_E)$. The states $(\mathcal{N} \otimes \mathcal{I}_E)(\rho'_{A'B'E})$ and $(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$ are related by

$$(\mathcal{N} \otimes \mathcal{I}_E)(\rho'_{A'B'E}) = (\mathbb{1}_{XYE} \otimes \sigma_1 \otimes \sigma_1)(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})(\mathbb{1}_{XYE} \otimes \sigma_1 \otimes \sigma_1)\,.$$

In other words $(\mathcal{N} \otimes \mathcal{I}_E)(\rho'_{A'B'E})$ is identical to $(\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$, except that the outcomes of each device have been relabelled $(a, b) \to (a \oplus 1, b \oplus 1)$. Note that our choice of the generalized CHSH score is invariant under the re-labelling of the outputs. It follows that $S((\mathcal{N} \otimes \mathcal{I}_E)(\rho'_{A'B'E})) = \omega$ and $\bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho'_{A'B'E})) = \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}))$. Now consider the state $\tilde{\rho}_{A'B'EE'} = (\rho_{A'B'E} \otimes |0\rangle\langle 0|_{E'} + \rho'_{A'B'E'} \otimes |1\rangle\langle 1|_{E'})/2$. We have $(\mathcal{N} \otimes \mathcal{I}_{EE'})(\tilde{\rho}_{A'B'EE'}) = ((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}) \otimes |0\rangle\langle 0|_{E'} + (\mathcal{N} \otimes \mathcal{I}_E)(\rho'_{A'B'E'}) \otimes |1\rangle\langle 1|_{E'})/2$. Since the score is linear, we have $S((\mathcal{N} \otimes \mathcal{I}_{EE'})(\tilde{\rho}_{A'B'EE'})) = \omega$. By construction, $\tilde{\rho}_{A'B'E} = (\sigma_2 \otimes \sigma_2 \otimes \mathbb{1}_E)\tilde{\rho}_{A'B'E}(\sigma_2 \otimes \sigma_2 \otimes \mathbb{1}_E)$. Finally, as a consequence of Lemma 10, for any of the entropy functions H we have $\bar{H}(\tilde{\rho}_{A'B'EE'}) = \bar{H}(\rho_{A'B'E})$. □

Corollary 2. *Any optimization of the form (4.9) is equivalent to an optimization of the same form but where each of the projectors are onto states of the form* $\alpha|0\rangle + \beta|1\rangle$ *with* $\alpha, \beta \in \mathbb{R}$ *and* $\rho_{A'B'E} = (\sigma_2 \otimes \sigma_2 \otimes \mathbb{1})\rho_{A'B'E}(\sigma_2 \otimes \sigma_2 \otimes \mathbb{1})$.

Next we consider the form of the reduced state $\rho_{A'B'}$ in the Bell basis.

Lemma 11. *Let* $\mathcal{N}$ *be the channel associated with a single-round CHSH strategy in which each POVM element is a projector of the form* $\cos(\alpha)|0\rangle + \sin(\alpha)|1\rangle$ *with* $\alpha \in \mathbb{R}$. *The state* $\rho^P_{A'B'E}$ *satisfies* $(\mathcal{N}\otimes\mathcal{I}_E)(\rho^P_{A'B'E}) = (\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E})$, *where* $\rho^P_{A'B'E}$ *is formed from* $\rho_{A'B'E}$ *by taking the partial transpose on* $A'B'$ *in the Bell basis.*

Proof. By definition, the partial transpose generates the state

$$\rho^P_{A'B'E} = \sum_{ij}(|\Psi_i\rangle\langle\Psi_j| \otimes \mathbb{1}_E)\rho_{A'B'E}(|\Psi_i\rangle\langle\Psi_j| \otimes \mathbb{1}_E)\,. \tag{4.10}$$

Writing the partial trace out in the Bell basis, for any two projectors Π_1 and Π_2 on $\mathcal{H}_{A'}$ and $\mathcal{H}_{B'}$ we have

$$\begin{aligned}
\mathrm{tr}\Big((\Pi_1 \otimes \Pi_2 \otimes \mathbb{1}_E)\rho^P_{A'B'E}\Big) &= \sum_i(\langle\Psi_i|(\Pi_1 \otimes \Pi_2) \otimes \mathbb{1}_E)\rho^P_{A'B'E}(|\Psi_i\rangle \otimes \mathbb{1}_E) \\
&= \sum_{ijk}((\langle\Psi_i|(\Pi_1 \otimes \Pi_2)|\Psi_j\rangle\langle\Psi_k|) \otimes \mathbb{1}_E)\rho_{A'B'E} \\
&\qquad (|\Psi_j\rangle\langle\Psi_k|\,|\Psi_i\rangle \otimes \mathbb{1}_E) \\
&= \sum_{ij}\langle\Psi_i|(\Pi_1 \otimes \Pi_2)|\Psi_j\rangle\,(\langle\Psi_i| \otimes \mathbb{1}_E)\rho_{A'B'E} \\
&\qquad (|\Psi_j\rangle \otimes \mathbb{1}_E)
\end{aligned} \tag{4.11}$$

When Π_1 and Π_2 are each projectors onto states of the form $\cos(\alpha)|0\rangle + \sin(\alpha)|1\rangle$ a short calculation reveals $\langle\Psi_i|(\Pi_1 \otimes \Pi_2)|\Psi_j\rangle = \langle\Psi_j|(\Pi_1 \otimes \Pi_2)|\Psi_i\rangle$. Using this in (4.11) we can conclude that

$$\mathrm{tr}_{A'B'}((\Pi_1 \otimes \Pi_2 \otimes \mathbb{1}_E)\rho^P_{A'B'E}) = \mathrm{tr}_{A'B'}((\Pi_1 \otimes \Pi_2 \otimes \mathbb{1}_E)\rho_{A'B'E})\,,$$

from which it follows that $(\mathcal{N} \otimes \mathcal{I}_E)(\rho^P_{A'B'E}) = (\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$. □

Corollary 3. *Any optimization of the form (4.9) is equivalent to an optimization of the same form but where each of the projectors are onto states of the form* $\alpha|0\rangle + \beta|1\rangle$ *with* $\alpha, \beta \in \mathbb{R}$, $\rho_{A'B'E} = (\sigma_2 \otimes \sigma_2 \otimes \mathbb{1})\rho_{A'B'E}(\sigma_2 \otimes \sigma_2 \otimes \mathbb{1})$ *and* $\rho_{A'B'E} = \rho^P_{A'B'E}$.

Proof. We established the invariance under $(\sigma_2 \otimes \sigma_2 \otimes \mathbb{1})$ in Corollary 2. Since $(\mathcal{N} \otimes \mathcal{I}_E)(\rho^P_{A'B'E}) = (\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$, if Eve uses the state $(\rho_{A'B'E} \otimes |0\rangle\langle 0|_{E'} + \rho^P_{A'B'E} \otimes |1\rangle\langle 1|_{E'})/2$, then, by the same argument used at the end of the proof of Lemma 7, the entropy and scores are unchanged while the state satisfies the required conditions. □

The next step is to show that the state on $A'B'$ can be taken to come from the set of density operators that are diagonal in the Bell basis. We define

$$\begin{aligned}\mathcal{S}_B := \{&\lambda_0 |\Phi_0\rangle\langle\Phi_0| + \lambda_1 |\Phi_1\rangle\langle\Phi_1| + \lambda_2 |\Phi_2\rangle\langle\Phi_2| + \lambda_3 |\Phi_3\rangle\langle\Phi_3| : 1 \geq \lambda_0 \geq \lambda_3 \geq 0, \\ &1 \geq \lambda_1 \geq \lambda_2 \geq 0, \lambda_0 - \lambda_3 \geq \lambda_1 - \lambda_2, \sum_i \lambda_i = 1\},\end{aligned} \tag{4.12}$$

where the states $\{|\phi_i\rangle\}_i$ are defined by (4.3)–(4.6).

Lemma 12. *Any optimization of the form (4.9) is equivalent to an optimization of the same form but where each of the projectors are onto states of the form* $\cos(\alpha)|0\rangle + \sin(\alpha)|1\rangle$ *with* $\alpha \in \mathbb{R}$ *and* $\rho_{A'B'} \in \mathcal{S}_B$.

Proof. From Corollary 3, we have that $\rho_{A'B'}$ can be taken to be invariant under $\sigma_2 \otimes \sigma_2$. Hence we can write

$$\rho_{A'B'} = \begin{pmatrix} \lambda'_0 & 0 & 0 & r_1 \\ 0 & \lambda'_1 & r_2 & 0 \\ 0 & r_2^* & \lambda'_2 & 0 \\ r_1^* & 0 & 0 & \lambda'_3 \end{pmatrix} \tag{4.13}$$

where the matrix is expressing the coefficients in the Bell basis. From corollary 3 we can impose that $\rho_{A'B'} = \rho^T_{A'B'}$, which then implies that r_1 and r_2 are real. Note that in order that $\rho_{A'B'}$ is a positive operator we require $r_1^2 \leq \lambda'_0\lambda'_3$ and $r_2^2 \leq \lambda'_1\lambda'_2$.
Let $U_\theta = \cos(\theta/2)|0\rangle\langle 0| + \sin(\theta/2)|0\rangle\langle 1| - \sin(\theta/2)|1\rangle\langle 0| + \cos(\theta/2)|1\rangle\langle 1|$, so that U_θ preserves the set $\{\cos(\alpha)|0\rangle + \sin(\alpha)|1\rangle : \alpha \in \mathbb{R}\}$. We proceed to show that for any state of the form (4.13) with r_1 and r_2 real, there exist values of θ_A and θ_B such that

$$\rho'_{A'B'} = (U_{\theta_A} \otimes U_{\theta_B})\rho_{A'B'}(U^\dagger_{\theta_A} \otimes U^\dagger_{\theta_B})$$

is diagonal in the Bell basis. We can compute the form of $\rho'_{A'B'}$ in the Bell basis. This has the same form as (4.13), but with r_1 replaced by $r_1\cos(\theta_A - \theta_B) + \frac{\lambda'_0 - \lambda'_3}{2}\sin(\theta_A - \theta_B)$ and r_2 replaced by $r_2\cos(\theta_A + \theta_B) + \frac{\lambda'_2 - \lambda'_1}{2}\sin(\theta_A + \theta_B)$. To make these zero we need

to choose θ_A and θ_B such that $\cos^2(\theta_A - \theta_B) = \frac{(\lambda'_0-\lambda'_3)^2}{(\lambda'_0-\lambda'_3)^2+4r_1^2}$ and $\cos^2(\theta_A + \theta_B) = \frac{(\lambda'_1-\lambda'_2)^2}{(\lambda'_1-\lambda'_2)^2+4r_2^2}$. If we write

$$\phi_1 = \cos^{-1}\left(\frac{\lambda'_0 - \lambda'_3}{\sqrt{(\lambda'_0 - \lambda'_3)^2 + 4r_1^2}}\right), \quad \phi_2 = \cos^{-1}\left(\frac{\lambda'_1 - \lambda'_2}{\sqrt{(\lambda'_1 - \lambda'_2)^2 + 4r_2^2}}\right),$$
$$\zeta_A = \frac{\phi_1 + \phi_2}{2} \quad \text{and} \quad \zeta_B = \frac{\phi_1 - \phi_2}{2}$$

then we can express the four solutions

$$(\theta_A, \theta_B) = (\zeta_A, \zeta_B),\ (\zeta_A + \pi/2, \zeta_B - \pi/2),\ (\zeta_A + \pi/2, \zeta_B + \pi/2),\ (\zeta_A + \pi, \zeta_B)\,.$$

Each of these brings the state into the form $\rho_{A'B'} = \lambda_0 |\Phi_0\rangle\langle\Phi_0| + \lambda_1 |\Phi_1\rangle\langle\Phi_1| + \lambda_2 |\Phi_2\rangle\langle\Phi_2| + \lambda_3 |\Phi_3\rangle\langle\Phi_3|$. The difference between the first two of these is an exchange of λ_0 with λ_3, the difference between the first and the third is an exchange of λ_1 with λ_2 and the difference between the first and the fourth is an exchange of λ_0 with λ_3 and of λ_1 with λ_2. It follows that we can ensure $\lambda_0 \geq \lambda_3$ and $\lambda_1 \geq \lambda_2$. Finally, if $\lambda_0 - \lambda_3 < \lambda_1 - \lambda_2$ we can apply $\sigma_3 \otimes \mathbb{1}$ to the resulting state, which simultaneously switches λ_0 with λ_1 and λ_2 with λ_3, while again preserving the set $\{\cos(\alpha)|0\rangle + \sin(\alpha)|1\rangle : \alpha \in \mathbb{R}\}$. □

The culmination of this section is the following.

Lemma 13. *For given p_{XY}, let*

$$\begin{aligned}
\mathcal{R}_1(\omega) &:= \{(\mathcal{N}, \rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_{\Pi_1,2}(p_{XY}),\\
&\qquad \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E),\ S((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \omega\} \quad \textit{and}\\
\mathcal{R}_2(\omega) &:= \{(\mathcal{N}, \rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_{\Pi_1,2}(p_{XY}),\\
&\qquad \rho_{A'B'E} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E),\ \rho_{A'B'} \in \mathcal{S}_B,\ S((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \omega\}\,.
\end{aligned}$$

We have $\inf_{\mathcal{R}_2(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \inf_{\mathcal{R}_1(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}))$ for any of the six entropy functions $\bar{H}$.

4.4 REDUCTION TO PURE STATES

Here we show that it is sufficient to restrict any of the optimizations we are interested in to pure states.

Lemma 14. *For given p_{XY} let $\mathcal{R}_2(\omega)$ be as in Lemma 13 and consider*

$$\begin{aligned}\mathcal{R}_3(\omega) \;:=\; \{(\mathcal{N}, \rho_{A'B'E}) : \mathcal{N} \in \mathcal{C}_{\Pi_1,2}(p_{XY}),\, \rho_{A'B'E} \in \mathcal{S}_P(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E),\\ \rho_{A'B'} \in \mathcal{S}_B,\, S((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \omega\}\,.\end{aligned}$$

We have $\inf_{\mathcal{R}_3(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \inf_{\mathcal{R}_2(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}))$ for any of the six entropy functions $\bar{H}$.

Proof. Since $\mathcal{R}_3 \subset \mathcal{R}_2$, we have $\inf_{\mathcal{R}_3(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) \geq \inf_{\mathcal{R}_2(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}))$. For the other direction, consider a state $\rho_{A'B'E}$ from the set $\mathcal{R}_2$, and let $\rho_{A'B'EE'}$ be its purification. Using the strong subadditivity of the von-Neumann entropy, $H(C|ZEE') \leq H(C|ZE)$, a new strategy in which the only change is that Eve holds a purification of $\rho_{A'B'E}$ cannot increase any of the entropic quantities of interest and makes no change to the score. Thus, $\inf_{\mathcal{R}_3(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) \leq \inf_{\mathcal{R}_2(\omega)} \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E}))$. □

Note that this also means that we can restrict $\mathcal{H}_E$ to be 4 dimensional.

4.5 SIMPLIFICATIONS OF QUBIT STRATEGIES FOR SPECIFIC ENTROPIC QUANTITIES

In this section, we compute expressions for each of the entropies of interest, based on the simplifications from the previous section. In other words, we are considering the optimizations

$$G(\omega, p_{XY}) := \min\ \bar{H}((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) \tag{4.14}$$

$$\mathcal{H}_{A'} = \mathcal{H}_{B'} = \mathbb{C}^2,\ \mathcal{H}_E = \mathbb{C}^4, \tag{4.15}$$

$$\rho_{A'B'E} \in \mathcal{S}_P(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E),\ \rho_{A'B'} \in \mathcal{S}_B$$

$$\mathcal{N}:\ \text{is of the form 4.20}$$

$$\left|\phi^A_{a|x}\right\rangle = \cos\left(\alpha_{a|x}\right)|0\rangle + \sin\left(\alpha_{a|x}\right)|1\rangle \quad \text{and} \tag{4.16}$$

$$\left|\phi^B_{b|y}\right\rangle = \cos\left(\beta_{b|y}\right)|0\rangle + \sin\left(\beta_{b|y}\right)|1\rangle \tag{4.17}$$

$$\alpha_{1|x} = \pi/2 + \alpha_{0|x} \quad \text{and} \quad \beta_{1|x} = \pi/2 + \beta_{0|x} \tag{4.18}$$

$$S((\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})) = \omega. \tag{4.19}$$

For convenience we sometimes use $\alpha_x = \alpha_{0|x}$ and $\beta_x = \beta_{0|x}$ and

$$\sigma_{A'B'} \mapsto \sum_{abxy} p_{XY}(x,y)\,|abxy\rangle\langle abxy|\,\mathrm{tr}\left(\left(\left|\phi^A_{a|x}\right\rangle\!\left\langle\phi^A_{a|x}\right| \otimes \left|\phi^B_{b|y}\right\rangle\!\left\langle\phi^B_{b|y}\right|\right)\sigma_{A'B'}\right) \quad (4.20)$$

Let

$$\begin{aligned}\tau_{ABXYE} &= (\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E}) \\ &= \sum_{abxy} p_{XY}(x,y)\,|abxy\rangle\langle abxy| \otimes \mathrm{tr}\left(\left(\left|\phi^A_{a|x}\right\rangle\!\left\langle\phi^A_{a|x}\right| \otimes \left|\phi^B_{b|y}\right\rangle\!\left\langle\phi^B_{b|y}\right| \otimes \mathbb{1}_E\right)\rho_{A'B'E}\right) \\ &= \sum_{abxy} p_{XY}(x,y)p_{AB|xy}(a,b)\,|abxy\rangle\langle abxy| \otimes \tau_E^{abxy}\,, \end{aligned} \quad (4.21)$$

where $\{\tau_E^{abxy}\}$ are normalized and tr is with respect to the systems $A'B'$.
We make a few initial observations.
Consider $p_{AB|xy}(a,b)\tau_E^{abxy} = \mathrm{tr}_{A'B'}\left(\left(\left|\phi^A_{a|x}\right\rangle\!\left\langle\phi^A_{a|x}\right| \otimes \left|\phi^B_{b|y}\right\rangle\!\left\langle\phi^B_{b|y}\right| \otimes \mathbb{1}_E\right)\rho_{A'B'E}\right)$. Since $\rho_{A'B'E}$ is pure, we can use the Schmidt decomposition to write $\rho_{A'B'E} = |\Phi\rangle\langle\Phi|_{A'B'E}$, where

$$|\Phi\rangle_{A'B'E} = \sum_i \sqrt{\lambda_i}\,|\Psi_i\rangle \otimes |i\rangle$$

where $\{|i\rangle\}$ is an orthonormal basis for $\mathcal{H}_E$. We have

$$\begin{aligned} p_{AB|xy}(a,b)\tau_E^{abxy} &= \sum_{ij}\sqrt{\lambda_i\lambda_j}\left(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|\right)|\Psi_i\rangle\,\langle\Psi_j|\left(\left|\phi^A_{a|x}\right\rangle \otimes \left|\phi^B_{b|y}\right\rangle\right)|i\rangle\langle j| \\ &= \left|\zeta^{abxy}\right\rangle\!\left\langle\zeta^{abxy}\right|, \end{aligned}$$

where

$$\left|\zeta^{abxy}\right\rangle = \sum_i \sqrt{\lambda_i}(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|)\,|\Psi_i\rangle\,|i\rangle \quad \text{and} \quad (4.22)$$

$$p_{AB|xy}(a,b) = \sum_i \lambda_i \left|(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|)\,|\Psi_i\rangle\right|^2. \quad (4.23)$$

Hence τ_E^{abxy} is pure for each a,b,x,y. Note also that

$$\left(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|\right)|\Psi_0\rangle = \frac{\cos\left(\beta_{b|y}-\alpha_{a|x}\right)}{\sqrt{2}} \quad (4.24)$$

$$\left(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|\right)|\Psi_1\rangle = \frac{\cos\left(\beta_{b|y}+\alpha_{a|x}\right)}{\sqrt{2}} \quad (4.25)$$

$$\left(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|\right)|\Psi_2\rangle = \frac{\sin\left(\beta_{b|y}+\alpha_{a|x}\right)}{\sqrt{2}} \quad (4.26)$$

$$\left(\left\langle\phi^A_{a|x}\right| \otimes \left\langle\phi^B_{b|y}\right|\right)|\Psi_3\rangle = \frac{\sin\left(\beta_{b|y}-\alpha_{a|x}\right)}{\sqrt{2}} \quad (4.27)$$

Because τ_{ABXYE} is formed from $\rho_{A'B'E}$ without acting on E, we have $H(E)_\tau = H(E)_\rho$, and because $\rho_{A'B'E}$ is pure, $H(E)_\rho = H(A'B')_\rho = H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\})$.[3] For the same reason, $\sum_{ab} p_{AB}(a,b)\tau_E^{abxy} = \tau_E$ for all x, y.

Lemma 15. *For $\sigma_{AXE} = \sum_{ax} p_{AX}(a,x)\,|a\rangle\langle a| \otimes |x\rangle\langle x| \otimes \sigma_E^{a,x}$, we have*

$$H(A|XE) = H(A|X) + \sum_{ax} p_{AX}(a,x) H(\sigma_E^{a,x}) - \sum_x p_X(x) H\left(\sum_a p_{A|x}(a)\sigma_E^{a,x}\right).$$

Proof. We have

$$\begin{aligned} H(A|XE) &= H(AXE) - H(XE) \\ &= H(AX) + \sum_{ax} p_{AX}(a,x) H(E|A=a, X=x) - H(X) - \\ &\qquad\qquad \sum_x p_X(x) H(E|X=x) \\ &= H(A|X) + \sum_{ax} p_{AX}(a,x)\left(H(\sigma_E^{a,x}) - H\left(\sum_{a'} p_{A|x}(a')\sigma_E^{a',x}\right)\right). \end{aligned}$$ □

We can parameterize the Bell diagonal state in the following way:

$$\lambda_0 = \frac{1}{4} + \frac{R\cos(\theta)}{2} + \delta \tag{4.28}$$

$$\lambda_1 = \frac{1}{4} + \frac{R\sin(\theta)}{2} - \delta \tag{4.29}$$

$$\lambda_2 = \frac{1}{4} - \frac{R\sin(\theta)}{2} - \delta \tag{4.30}$$

$$\lambda_3 = \frac{1}{4} - \frac{R\cos(\theta)}{2} + \delta \tag{4.31}$$

where $0 \le R \le 1$, $0 \le \theta \le \pi/4$ if $R \le 1/\sqrt{2}$, or $0 \le \theta \le \pi/4 - \cos^{-1}(1/(R\sqrt{2}))$ if $R > 1/\sqrt{2}$ and $-1/4 + R\cos(\theta)/2 \le \delta \le 1/4 - R\sin(\theta)/2$.

Lemma 16. *For $R > 1/\sqrt{2}$, $\max_\delta H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\})$ is achieved when $\delta = \delta^* = \frac{R^2\cos(2\theta)}{4}$.*

Proof. One can compute the derivative of $H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\})$ with respect to δ to see that it is 0 only for $\delta = \delta^* := \frac{R^2\cos(2\theta)}{4}$.

We next check that δ^* is in the valid range of δ. The condition $\delta^* \le 1/4 - R\sin(\theta)/2$ rearranges to $2R^2\sin^2(\theta) - 2R\sin(\theta) + 1 - R^2 \ge 0$. The roots of the quadratic equation $2R^2x^2 - 2Rx + 1 - R^2$ are at $x = \frac{1}{2R}(1 \pm \sqrt{2R^2 - 1})$. For $R \ge \frac{1}{\sqrt{2}}$ the roots are real[4]. Our condition on θ implies that $0 \le \sin(\theta) \le \frac{1}{2R}(1 - \sqrt{2R^2 - 1})$, hence taking

[3] We use H for both the von Neumann and Shannon entropies; if a list of probabilities is given as the argument to H it signifies the Shannon entropy.

[4] If $R < \frac{1}{\sqrt{2}}$ there are no real roots and the condition always holds.

$x = \sin(\theta)$ we are always to the left of the first root and so $\delta^* \leq 1/4 - R\sin(\theta)/2$. A similar argument shows $\delta^* \geq -1/4 + R\cos(\theta)/2$.
We can then compute the double derivative of $H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\})$ with respect to δ and evaluate it at δ^*. This gives $-\frac{32}{\ln(2)(2-4R^2+R^4+R^4\cos(4\theta))}$, which can be shown to be negative for $R > 1/\sqrt{2}$ and any valid θ using a similar argument to that above. □

Using this state and the specified measurements, the probability table for the observed distribution has the form given in the table below (whose entries correspond to $p_{AB|XY}$):

		$Y=0$		$Y=1$	
		$B=0$	$B=1$	$B=0$	$A=1$
$X=0$	$A=0$	ϵ_{00}	$\frac{1}{2}-\epsilon_{00}$	ϵ_{01}	$\frac{1}{2}-\epsilon_{01}$
	$A=1$	$\frac{1}{2}-\epsilon_{00}$	ϵ_{00}	$\frac{1}{2}-\epsilon_{01}$	ϵ_{01}
$X=1$	$A=0$	ϵ_{10}	$\frac{1}{2}-\epsilon_{10}$	$\frac{1}{2}-\epsilon_{11}$	ϵ_{11}
	$A=1$	$\frac{1}{2}-\epsilon_{10}$	ϵ_{10}	ϵ_{11}	$\frac{1}{2}-\epsilon_{11}$

where

$$\epsilon_{00} = \frac{1}{4}\left(1 + R\cos(\theta)\cos(2(\alpha_0 - \beta_0)) + R\sin(\theta)\cos(2(\alpha_0 + \beta_0))\right)$$
$$\epsilon_{01} = \frac{1}{4}\left(1 + R\cos(\theta)\cos(2(\alpha_0 - \beta_1)) + R\sin(\theta)\cos(2(\alpha_0 + \beta_1))\right)$$
$$\epsilon_{10} = \frac{1}{4}\left(1 + R\cos(\theta)\cos(2(\alpha_1 - \beta_0)) + R\sin(\theta)\cos(2(\alpha_1 + \beta_0))\right)$$
$$\epsilon_{11} = \frac{1}{4}\left(1 - R\cos(\theta)\cos(2(\alpha_1 - \beta_1)) - R\sin(\theta)\cos(2(\alpha_1 + \beta_1))\right) .$$

Note that

$$S(\tau_{ABXY}) = 2\sum_{ij}\gamma_{ij}\epsilon_{ij} \tag{4.32}$$

is independent of δ. In the upcoming sections, we use these simplifications to obtain analytic forms for entropic quantities for the qubit case.

4.6 H(AB|X=0,Y=0,E)

For $H(AB|X=0,Y=0,E)$ we are interested in the state

$$\tau'_{ABE} = \sum_{ab} p_{AB|00}(a,b)\,|a\rangle\langle a| \otimes |b\rangle\langle b| \otimes \tau_E^{ab00}$$

since $H(AB|X=0,Y=0,E)_\tau = H(AB|E)_{\tau'}$. Note that, as above, $H(E)_{\tau'} = H(E)_\rho$. Using Lemma 15 we have

$$\begin{aligned} H(AB|E)_{\tau'} &= H(AB)_{\tau'} + \sum_{ab} p_{AB}(a,b) H(\tau_E^{ab00}) - H\left(\sum_{ab} p_{AB}(a,b)\tau_E^{ab00}\right) \\ &= H(AB)_{\tau'} + \sum_{ab} p_{AB}(a,b) H(\tau_E^{ab00}) - H(E)_{\tau'} \,. \end{aligned}$$

However, since τ_E^{ab00} is pure for each a, b, $H(\tau_E^{ab00}) = 0$ and we find

$$\begin{aligned} H(AB|E)_{\tau'} &= H(AB)_{\tau'} - H(E)_\rho \\ &= H(\{\epsilon_{00}, \epsilon_{00}, 1/2-\epsilon_{00}, 1/2-\epsilon_{00}\}) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\}) \\ &= 1 + H_{\text{bin}}(2\epsilon_{00}) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\}) \,. \end{aligned}$$

Lemma 16 shows that $\max_\delta H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\})$ is achieved for $\delta = \delta^* = \frac{R^2\cos(2\theta)}{4}$. Since the score is independent of δ we can take the state to satisfy $\delta = \delta^*$ and remove δ from the optimization.

4.7 H(AB|XYE)

In this case we again use Lemma 15 to obtain

$$\begin{aligned} H(AB|XYE) &= H(AB|XY) + \sum_{abxy} p_{ABXY}(a,b,x,y) H(\tau_E^{abxy}) - \\ &\qquad\qquad \sum_{xy} p_{XY}(x,y) H\left(\sum_{ab} p_{AB|xy}(a,b)\tau_E^{abxy}\right) \\ &= H(AB|XY) - H(E) \,, \end{aligned}$$

where we again use that $H(\tau_E^{abxy}) = 0$, and note that $\sum_{ab} p_{AB|xy}(a,b)\tau_E^{abxy} = \rho_E$ for all x, y. Note that

$$\begin{aligned} H(AB|XY) &= \sum_{xy} p_{XY}(x,y) H(AB|X=x,Y=y) \\ &= 1 + \sum_{xy} p_{XY}(x,y) H_{\text{bin}}(2\epsilon_{xy}) \,, \end{aligned}$$

and so we have

$$H(AB|XYE) = 1 + \sum_{xy} p_{XY}(x,y) H_{\text{bin}}(2\epsilon_{xy}) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\}) \,. \tag{4.33}$$

This is again independent of δ, so, like in the case of $H(AB|X=0,Y=0,E)$ we can take $\delta = \delta^*$ and remove δ from the optimization.

4.8 H(AB|E)

We first trace out XY to give $\tau_{ABE} = \sum_{ab} p_{AB} \left|a\right\rangle\!\left\langle a\right| \otimes \left|b\right\rangle\!\left\langle b\right| \otimes \sum_{xy} p_{XY|ab}(x,y)\tau_E^{abxy}$. For this state, Lemma 15 gives

$$\begin{aligned} H(AB|E) &= H(AB) + \sum_{ab} p_{AB}(a,b) H\left(\sum_{xy} p_{XY|ab}(x,y)\tau_E^{abxy}\right) - \\ &\qquad\qquad H\left(\sum_{abxy} p_{AB}(a,b)p_{XY|ab}(x,y)\tau_E^{abxy}\right) \\ &= H(AB) + \sum_{ab} p_{AB}(a,b) H\left(\sum_{xy} p_{XY|ab}(x,y)\tau_E^{abxy}\right) - H(E) \\ &= H(AB) + \sum_{ab} p_{AB}(a,b) H\left(\sum_{xy} \frac{p_{XY}(x,y)p_{AB|xy}(a,b)\tau_E^{abxy}}{p_{AB}(a,b)}\right) - H(E)\,. \end{aligned}$$

In this case we cannot remove the middle term, and the middle term is not independent of δ. The optimization in this case is hence significantly more complicated. Note that

$$H(AB) = 1 + H_{\text{bin}}\left(2\left(\sum_{(x,y)\neq(1,1)} p_{XY}(x,y)\epsilon_{xy} + p_{XY}(1,1)\left(\frac{1}{2} - \epsilon_{11}\right)\right)\right)\,.$$

4.9 H(A|X=0,Y=0,E)

For this section, we only focus on the case when ω is the CHSH score. This case was already covered in [42] where it was solved analytically (see also [45] for a slight generalization).

Lemma 17. *For $3/4 \leq \omega \leq \frac{1}{2}(1+\frac{1}{\sqrt{2}})$ the solution to the optimization problem (4.14) when $\bar{H} = H(A|X=0,Y=0,E)$ is $1 - H_{\text{bin}}\left(\frac{1}{2}(1+\sqrt{16\omega(\omega-1)+3})\right)$.*

For completeness we give a proof here as well. We first show that the maximum CHSH score for a Bell diagonal state depends only on R.

Lemma 18. *Given a state $\rho_{A'B'E}$ with $\rho_{A'B'}$ parameterized as in (4.28)–(4.31), if $\mathcal{N}$ satisfies the requirements of the optimization problem (4.14), then $\tau_{ABXYE} = (\mathcal{N} \otimes \mathcal{I}_E)(\rho_{A'B'E})$ satisfies $S(\tau_{ABXYE}) \leq \frac{1}{2} + \frac{R}{2\sqrt{2}}$, and there exists a channel $\mathcal{N}$ achieving equality.*

Proof. Consider the score function (4.32). Collecting all the terms involving α_0 and α_1 and some manipulation gives

$$S(\tau_{ABXYE}) = \frac{1}{2} + \frac{R}{2\sqrt{2}}\cos(\beta_0-\beta_1)A_1 + \frac{R}{2\sqrt{2}}\sin(\beta_0-\beta_1)B_1,$$

where we have used $\cos(\theta)+\sin(\theta) = \sqrt{2}\sin\left(\frac{\pi}{4}+\theta\right)$ and $\cos(\theta)-\sin(\theta) = \sqrt{2}\cos\left(\frac{\pi}{4}+\theta\right)$ and

$$A_1 = \left[\sin(2\alpha_0)\sin(\beta_0+\beta_1)\cos\left(\frac{\pi}{4}+\theta\right) + \cos(2\alpha_0)\cos(\beta_0+\beta_1)\sin\left(\frac{\pi}{4}+\theta\right)\right]$$
$$B_1 = \left[\sin(2\alpha_1)\cos(\beta_0+\beta_1)\cos\left(\frac{\pi}{4}+\theta\right) - \cos(2\alpha_1)\sin(\beta_0+\beta_1)\sin\left(\frac{\pi}{4}+\theta\right)\right]$$

. For brevity we write $\bar{\theta} = \frac{\pi}{4}+\theta$. We then use that for $r,t,\phi \in \mathbb{R}$ we have $r\cos(\phi) + t\sin(\phi) \leq \sqrt{r^2+t^2}$ with equality if $r\cos(\phi)+t\sin(\phi) \geq 0$ and $r\sin(\phi) = t\cos(\phi)$. This allows us to form the bound

$$\begin{aligned} S(\tau_{ABXYE}) \leq & \frac{1}{2} + \frac{R}{2\sqrt{2}}\left(|\cos(\beta_0-\beta_1)|\sqrt{\sin^2(\beta_0+\beta_1)\cos^2(\bar{\theta}) + \cos^2(\beta_0+\beta_1)\sin^2(\bar{\theta})}\right. \\ & \left. + |\sin(\beta_0-\beta_1)|\sqrt{\cos^2(\beta_0+\beta_1)\cos^2(\bar{\theta}) + \sin^2(\beta_0+\beta_1)\sin^2(\bar{\theta})}\right) \\ \leq & \frac{1}{2} + \frac{R}{2\sqrt{2}}\,. \end{aligned}$$

Choosing $\alpha_0 = 0$, $\alpha_1 = \pi/4$, $\beta_0 = \frac{\pi}{8} - \frac{\theta}{2}$, $\beta_1 = -\frac{\pi}{8}+\frac{\theta}{2}$ achieves equality (for instance). □

It follows that $\omega > 3/4$ is only possible if $R > 1/\sqrt{2}$.

We now turn to the entropy. In this case we trace out B from the state τ' in Section 4.6 to give $\tau'_{AE} = \sum_a p_{A|00}(a)\,|a\rangle\langle a| \otimes \sum_b p_{B|a00}(b)\tau_E^{ab00}$, so that $H(A|X=0,Y=0,E)_\tau = H(A|E)_{\tau'}$. Using Lemma 15 we have

$$\begin{aligned} H(A|E)_{\tau'} &= H(A)_{\tau'} + \sum_a p_{A|00}(a) H\left(\sum_b p_{B|a00}(b)\tau_E^{ab00}\right) - H\left(\sum_{ab} p_{AB|00}(a,b)\tau_E^{ab00}\right) \\ &= 1 + \sum_a \frac{1}{2} H\left(\sum_b p_{B|a00}(b)\tau_E^{ab00}\right) - H(E)_\rho \\ &= 1 + \sum_a \frac{1}{2} H\left(\sum_b 2p_{AB|00}(a,b)\tau_E^{ab00}\right) - H(E)_\rho\,, \end{aligned}$$

where we have used the fact that $p_{A|00}(a) = 1/2$ for $a = 0,1$. The eigenvalues of $\sum_b 2p_{AB|00}(a,b)\tau_E^{ab00}$ turn out to be

$$\frac{1}{2}\left(1 \pm \sqrt{2(\lambda_0-\lambda_3)(\lambda_1-\lambda_2)\cos(4\alpha_0) + (\lambda_0-\lambda_3)^2 + (\lambda_1-\lambda_2)^2}\right)\,,$$

independently of a. Hence, we can write $H(A|E)_{\tau'}$ in terms of the Bell diagonal state using

$$\sum_a \frac{1}{2} H\left(\sum_b 2p_{AB|00}(a,b)\tau_E^{ab00}\right) = \Phi_s\Big(2(\lambda_0-\lambda_3)(\lambda_1-\lambda_2)\cos(4\alpha_0) + (\lambda_0-\lambda_3)^2+(\lambda_1-\lambda_2)^2\Big)$$

$$H(E)_\rho = H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\})$$

where $\Phi_s(x) = H_{\text{bin}}(\frac{1}{2}+\frac{\sqrt{x}}{2})$ is a short-hand notation. Having established this, we show the following.

Lemma 19. *Let $\rho_{A'B'E}$ be pure with $\mathcal{H}_{A'} = \mathcal{H}_{B'} = \mathbb{C}^2$, and let $\rho_{A'B'}$ be a Bell diagonal state parameterized by (4.28)–(4.31) with $R > 1/\sqrt{2}$. Let τ be the state defined by (4.21) $H(A|X=0,Y=0,E)_\tau \geq 1 + H_{\text{bin}}\left(\frac{1}{2}\left(1+\sqrt{2R^2-1}\right)\right)$ where equality is achievable for $\alpha_0 = 0$.*

Proof. We note that

$$\sum_a \frac{1}{2} H\left(\sum_b 2p_{AB|00}(a,b)\tau_E^{ab00}\right) \geq \Phi_s\left(2(\lambda_0-\lambda_3)(\lambda_1-\lambda_2)+(\lambda_0-\lambda_3)^2+(\lambda_1-\lambda_2)^2\right)$$
$$= H_{\text{bin}}(\lambda_0+\lambda_1)\,,$$

with equality for $\alpha_0 = 0$. Hence

$$H(A|E)_{\tau'} \geq 1 + H_{\text{bin}}(\lambda_0+\lambda_1) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\})\,. \tag{4.34}$$

Using the parameterization of (4.28)–(4.31) we have $\lambda_0+\lambda_1 = 1/2(1+R(\cos(\theta)+\sin(\theta)))$. Thus, the minimum of $H(A|E)_{\tau'}$ over δ is achieved for $\delta = \delta^*$ (as in the case $H(AB|X=0,Y=0,E)$). Taking $\delta = \delta^*$ and differentiating the resulting expression with respect to θ yields

$$\frac{R}{2}(\cos(\theta)+\sin(\theta))\log\left(\frac{1-R\cos(\theta)+R\sin(\theta)}{1+R\cos(\theta)-R\sin(\theta)}\right)\,.$$

Since $\cos(\theta)+\sin(\theta) = \sqrt{2}\sin(\pi/4+\theta)$, the leading factor is always positive over our range of θ. The logarithm term is always negative, except for $\theta = \pi/4$ where it reaches zero. Thus, the minimum over θ is always obtained at the largest possible θ, i.e., $\theta = \pi/4 - \cos^{-1}\left(1/(R\sqrt{2})\right)$.
With this substitution the right hand side of (4.34) reduces to

$$1 + H_{\text{bin}}\left(\frac{1}{2}\left(1+\sqrt{2R^2-1}\right)\right)\,,$$

establishing the claim. □

Lemma 17 is then a corollary of Lemmas 18 and 19.

Proof of Lemma 17. From Lemma 19 we have

$$H(A|X=0,Y=0,E)_\tau \geq 1 + H_{\text{bin}}\left(\frac{1}{2}\left(1+\sqrt{2R^2-1}\right)\right) .$$

However, Lemma 18 then implies

$$\begin{aligned} H(A|X=0,Y=0,E)_\tau &\geq 1 + H_{\text{bin}}\left(\frac{1}{2}\left(1+\sqrt{4(2\omega-1)^2-1}\right)\right) \\ &= 1 + H_{\text{bin}}\left(\frac{1}{2}\left(1+\sqrt{16\omega(\omega-1)+3)}\right)\right) , \end{aligned}$$

where we use the fact that $H_{\text{bin}}(p)$ is decreasing and concave for $p \geq 1/2$. Equality is achievable by taking $\alpha_0 = 0$, $\alpha_1 = \pi/4$, $\beta_0 = \frac{\pi}{8} - \frac{\theta}{2}$, $\beta_1 = -\frac{\pi}{8} + \frac{\theta}{2}$. □

We use this case to gain confidence in our numerics, since we can make a direct comparison to the analytic curve.

4.10 H(A|XYE)

In this case Lemma 15 gives

$$\begin{aligned} H(A|XYE) &= H(A|XY) + \sum_{axy} p_{AXY}(a,x,y) H\left(\sum_b p_{B|axy}(b)\tau_E^{abxy}\right) \\ &\quad - \sum_{xy} p_{XY}(x,y) H\left(\sum_{ab} p_{AB|xy}(a,b)\tau_E^{abxy}\right) \\ &= 1 + \sum_{axy} p_{XY}(x,y) p_{A|xy}(a) H\left(\sum_b 2p_{AB|xy}(a,b)\tau_E^{abxy}\right) - H(E) . \end{aligned}$$

The eigenvalues of $\sum_b 2p_{AB|xy}(a,b)\tau_E^{abxy}$ can be computed to be

$$\frac{1}{2}\left(1 \pm \sqrt{2(\lambda_0-\lambda_3)(\lambda_1-\lambda_2)\cos(4\alpha_x) + (\lambda_0-\lambda_3)^2 + (\lambda_1-\lambda_2)^2}\right) ,$$

independently of a, y. If we define

$$g(\alpha) := \frac{1}{2}\left(1 + \sqrt{2(\lambda_0-\lambda_3)(\lambda_1-\lambda_2)\cos(4\alpha) + (\lambda_0-\lambda_3)^2 + (\lambda_1-\lambda_2)^2}\right)$$

then

$$H(A|XYE) = 1 + \textstyle\sum_x p_X(x) H_{\text{bin}}(g(\alpha_x)) - H(E) . \tag{4.35}$$

Note that using the parameterization (4.28)–(4.31) we have

$$g(\alpha) = \frac{1}{2}\left(1 + R\sqrt{1 + \sin(2\theta)\cos(4\alpha)}\right) .$$

Since this is independent of δ, we can again use $\delta = \delta^*$ to remove one parameter when minimizing $H(A|XYE)$.

4.11 H(A|E)

In this case Lemma 15 gives

$$\begin{aligned} H(A|E) &= H(A) + \sum_a p_A(a) H\left(\sum_{bxy} p_{BXY|a}(b,x,y)\tau_E^{abxy}\right) - \\ &\qquad\qquad H\left(\sum_{abxy} p_{AB}(a,b)p_{XY|ab}(x,y)\tau_E^{abxy}\right) \\ &= 1 + \frac{1}{2}\sum_a H\left(\sum_{bxy} 2p_{ABXY}(a,b,x,y)\tau_E^{abxy}\right) - H(E) \\ &= 1 + \frac{1}{2}\sum_a H\left(\sum_{bxy} 2p_{XY}(x,y)P_{AB|xy}(a,b)\tau_E^{abxy}\right) - H(E) \\ &= 1 + \frac{1}{2}\sum_a H\left(\sum_x 2p_X(x)\,\mathrm{tr}_{A'}\left(\left(\left|\phi^A_{a|x}\right\rangle\!\left\langle\phi^A_{a|x}\right| \otimes \mathbb{1}_E\right)\rho_{A'E}\right)\right) - H(E) . \end{aligned}$$

Like in the case $H(AB|E)$ the middle term cannot be removed and this term is not independent of δ.

4.12 NUMERICALLY COMPUTING UPPER BOUNDS ON RATES

As we have shown, the optimizations that define the G functions can be expressed in terms of at-most 7 real parameters (3 to specify the state and 4 to choose the measurements). They are hence amenable to numerical optimizations. We note also that except in the cases $G_{AB|E}(\omega, p_{XY})$ and $G_{A|E}(\omega, p_{XY})$ we can remove an additional parameter. In this section, we present a heuristic method to estimate the rates. We restrict ourselves to the case when ω is the CHSH score (i.e. $\gamma_{ij} = \frac{1}{4}$). This is also the case for the next chapter.

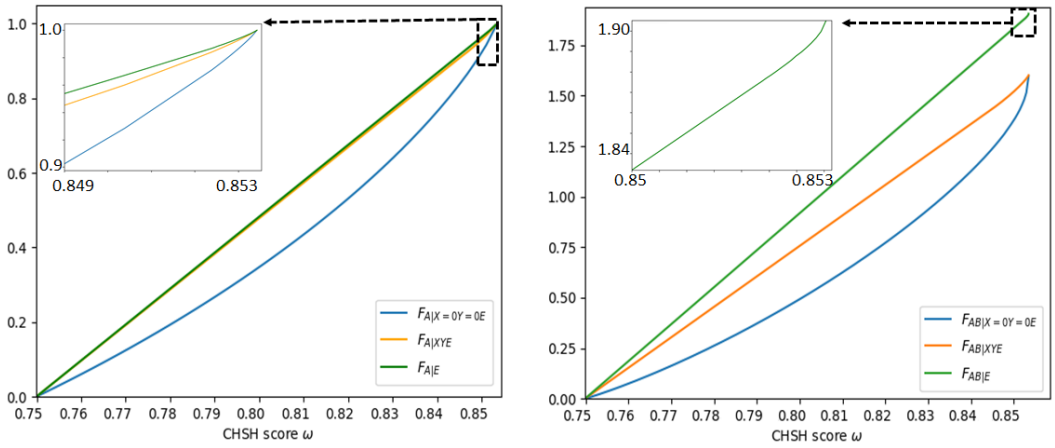
We obtain upper bounds by using numerical solvers that attempt to compute G (these give upper bounds because the computations are not guaranteed to converge). Our program for computing G runs in N iterations. In each iteration the program starts by making a random guess for the parameters from within the valid range. It then uses sequential quadratic programming to minimize $\bar{H}((\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E}))$ subject to the CHSH score being fixed [here $\bar{H}$ is a placeholder for one of the entropic quantities of interest]. On each iteration, the program arrives at a candidate for the minimum value, and we run $N \approx 10^4$ iterations to arrive at the conjectured minimum value for $\bar{H}((\mathcal{N}\otimes\mathcal{I}_E)(\rho_{A'B'E}))$. The numerical optimization is performed in Python using the sequential least squares programming (SLSQP) solver in SciPy. The curves obtained were found to match those generated by solving numerically in Mathematica and Matlab. Since these optimizations are not guaranteed to converge, the generated curves are upper bounds on the infima. Some confidence of their tightness comes from the smoothness of the curves, the consistency across different numerical solvers, and that the generated points match the known analytic tight bound in the case $H(A|X=0,Y=0,E)$. They also closely match the numerical lower bounds we computed for $G_{A|XYE}$ and $G_{AB|X=0,Y=0,E}$ discussed in the next chapter 5.

$G_{AB|00E}$ and $G_{A|00E}$ are convex functions, and hence $F=G$ for these. For the other cases we generate the graphs in the case where p_{XY} is uniform, observing that each of the G curves starts with a concave part and switches to convex for larger CHSH scores. Since the minimum entropy is always zero for classical scores, each of the G curves approach 0 as ω approaches $3/4$. Each of the F curves can be found from G by finding the tangent to G that passes through $(3/4,0)$. We call the score at which this tangent is taken ω^*, defined by $(\omega^*-3/4)G'(\omega^*)=G(\omega^*)$. We then have

$$F(\omega)=\begin{cases} G'(\omega^*)(\omega-\frac{3}{4}) & \text{if } \omega\leq\omega^* \\ G(\omega) & \text{otherwise} \end{cases}. \tag{4.36}$$

We give estimates for ω^* for each of the cases below. In essence, what this means is that for $\omega<\omega^*$ the optimal strategy for Eve is to either use a deterministic classical strategy with score $3/4$ or a strategy that achieves score ω^*, mixing these such that the average score is ω. Eve can remember which strategy she used, and hence the entropy from her perspective is also the convex mixture of the entropies of the endpoints.

Figure 4.1 shows the curves we obtained for the functions F in each of the six cases. Note that, except in the cases where we condition on $X=0$ and $Y=0$, the graphs all have linear sections as a result of taking the convex lower bound. In the figure 4.1 we show



(a) (b)

Figure 4.1: Graphs of the rates for (a) the one-sided and (b) the two-sided randomness with uniformly chosen inputs. Each of these curves has a non-linear part and the blue curves do not have a linear part.

the graphs for G together with those for F. The approximate coordinate of the top of the linear segment for $F_{AB|E}$ is $(0.8523, 1.8735)$ and for $F_{A|E}$ it is $(0.8505, 0.967)$. Note also that $1 + H_{\text{bin}}\left(\frac{1}{2} + \frac{1}{\sqrt{32}}\right) \approx 1.908$ is the maximum value on the graph $F_{AB|E}(\omega)$. By examining the parameters that come out of the numerical optimizations we have the following.

Lemma 20. *Consider the curve* $g_1(\omega) = 1 + H_{\text{bin}}(\omega) - 2H_{\text{bin}}(\frac{1}{2} + \frac{2\omega-1}{\sqrt{2}})$. $F_{AB|XYE}(\omega)$ *can be upper bounded in terms of* g_1 *as follows*

$$F_{AB|XYE}(\omega) \leq \begin{cases} g_1(\omega) & \omega^*_{AB|XYE} \leq \omega \leq \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right) \\ g_1'(\omega^*_{AB|XYE})(\omega - 3/4) & 3/4 \leq \omega \leq \omega^*_{AB|XYE} \end{cases}. \tag{4.37}$$

where $\omega^*_{AB|XYE} \approx 0.84403$ *is the solution to* $g_1'(\omega)(\omega - 3/4) = g_1(\omega)$. *Note that* $g_1(\omega^*_{AB|XYE}) \approx 1.4186$ *and the maximum value reached is* $1 + H_{\text{bin}}(1/2 + 1/(2\sqrt{2})) \approx 1.601$.

Proof. We first consider an upper bound on $G_{AB|XYE}(\omega)$. In section 4.5 we give a parameterization of a two-qubit state (with parameters R, θ and δ) and measurements (with parameters α_0, α_1, β_0 and β_1) before computing an expression for $H(AB|XYE)$ in terms of these (see (4.33)). We also obtain an expression for the CHSH score (see (4.32)).

Choosing $R = \sqrt{2}(2\omega - 1)$, $\theta = 0$, $\delta = R^2/4$, $\alpha_0 = 0$, $\alpha_1 = \pi/4$, $\beta_0 = \pi/8$, $\beta_1 = -\pi/8$ we find a score ω, and calculating $H(AB|XYE)$ we obtain $H(AB|XYE) = g_1(\omega)$ and hence $G_{AB|XYE}(\omega) \leq g_1(\omega)$. Since, $G_{AB|XYE}(3/4) = 0$, and $F_{AB|XYE}$ is formed from $G_{AB|XYE}$ by taking the convex lower bound, we establish the claim. □

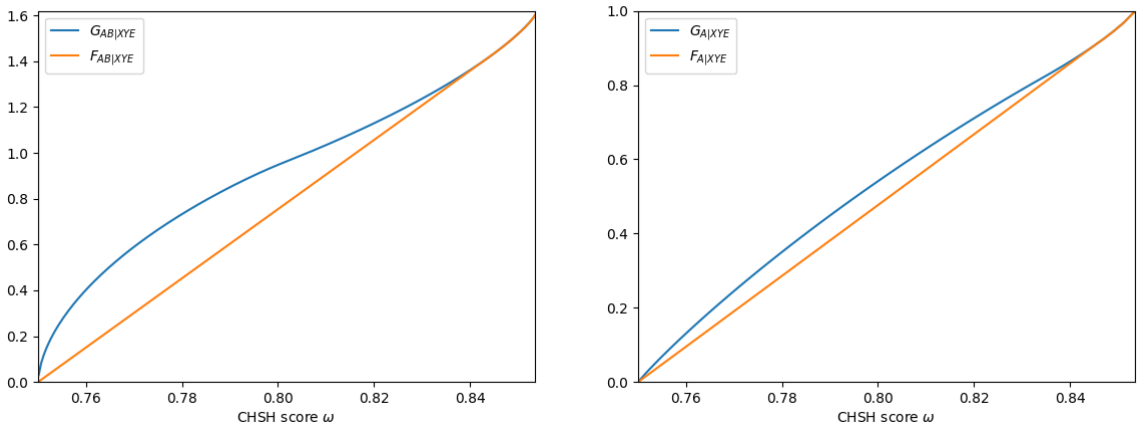
Lemma 21. *Consider the curve* $g_2(\omega) = 1 - H_{\text{bin}}\left(\frac{1}{2} + \frac{2\omega-1}{\sqrt{2}}\right)$. $F_{A|XYE}(\omega)$ *is upper bounded by the convex lower bound of* $g_2(\omega)$. *In other words,*

$$F_{A|XYE}(\omega) \leq \begin{cases} g_2(\omega) & \omega^*_{A|XYE} \leq \omega \leq \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right) \\ g_2'(\omega^*_{A|XYE})(\omega - 3/4) & 3/4 \leq \omega \leq \omega^*_{A|XYE} \end{cases}. \tag{4.38}$$

where $\omega^*_{A|XYE} \approx 0.84698$ *is the solution to* $g_2'(\omega)(\omega - 3/4) = g_2(\omega)$. *Note that* $g_2(\omega^*_{A|XYE}) \approx 0.92394$.

Proof. The proof is the same as for Lemma 20, except that g_1 is replaced by g_2 — the choice of state and measurements remains the same. □

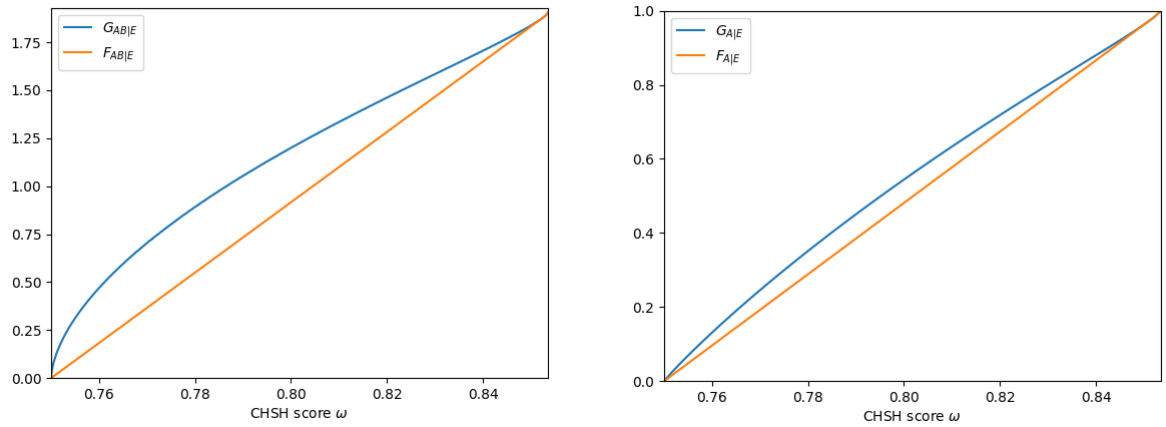
4.13 UPPER BOUNDS FOR OTHER ENTROPIC QUANTITIES



(a) (b)

Figure 4.2: (a) Two-sided and (b) one-sided entropy curves conditioned on X, Y and E with uniform input distribution.

Figure 4.2 gives one-sided and two-sided randomness rates as a function of CHSH score when inputs are chosen uniformly at random. Recall that the curves for $F_{A|E}(\omega)$ and



(a) (b)

Figure 4.3: (a) Two-sided and (b) one-sided entropy curves conditioned on E with uniform input distribution.

$F_{AB|E}(\omega)$ (Fig 4.3) are fundamental upper bounds on one-sided and two-sided output randomness rates as a function of the CHSH score. The curves $F_{.|.E}$ are obtained by taking the convex lower bound (of convex envelope) of $G_{.|.E}$, which is the randomness rates if only qubit-strategies were used – i.e. Alice and Bob are only allowed to share qubit strategies.

Figure 4.2(a) gives the graphs of $F_{AB|XYE}(\omega)$ and $G_{AB|XYE}(\omega)$, while Figure 4.2(b) shows those for $F_{A|XYE}(\omega)$ and $G_{A|XYE}(\omega)$. In each case the G graphs have a concave and convex part and the F graphs are formed by taking the convex lower bound. For these cases the points at which the tangents are taken are $\omega^*_{AB|XYE} \approx 0.8440$ and $\omega^*_{A|XYE} \approx 0.8470$.

Figure 4.3(a) gives the graphs of $F_{AB|E}(\omega)$ and $G_{AB|E}(\omega)$, while Figure 4.3(b) shows those for $F_{A|XYE}(\omega)$ and $G_{A|XYE}(\omega)$. Again, in each case the G graphs have a concave and convex part and the F graphs are formed by taking the convex lower bound. For these cases the points at which the tangents are taken are $\omega^*_{A|E} \approx 0.8505$ and $\omega^*_{AB|E} \approx 0.8523$. These curves are generated using heuristic numerical optimizations, and therefore can be only be treated as good estimates for the randomness rate. In the next chapter, we shall derive some of these curves using reliable and rigorous techniques.

— 5 —

Lower bounds on the entropies

5.1 LOWER BOUNDS

Lemma 21 gives an upper bound on the one-sided randomness using an explicit strategy. However, for security proofs a lower bound is needed. In this chapter, we compute such lower bounds for $G_{A|XYE}$, $G_{AB|X=0,Y=0,E}$ and $G_{AB|XYE}$. Note that for this chapter, we restrict to the case when ω is the CHSH score instead of the generalized CHSH score.
The idea behind our lower bounds is as follows. We first show that for every fixed value of ω the functions $G_{A|XYE}(\omega)$ and $G_{AB|X=0,Y=0,E}(\omega)$ can each be expressed as a minimization over 3 real parameters. For the function $G_{AB|XYE}$, we use polynomial optimization techniques to form a lower bound. For fixed ω we compute the values of the objective function on a grid of points comprising these parameters. By bounding the derivative of the objective function within the cuboids generated by the grid we establish a lower bound on the function over the possible parameters. The lower bound we generate can in principle be made arbitrarily good by decreasing the grid spacing (at the expense of taking more time to evaluate).
Given lower bounds on $G_{A|XYE}(\omega), G_{AB|X=0,Y=0,E}(\omega)$ and $G_{AB|XYE}(\omega)$ for a finite set of values of ω, we can get lower bounds for all values of ω by using that the G functions are monotonically increasing in ω, so we have $G(\omega) \leq G(\omega + \nu)$, where ν is the spacing between the finite set of values of ω. Hence, we can only compute the lower bounds for finitely many values $\mathcal{W} = \{\omega_1, \omega_2, \ldots\}$ in the range $(3/4, (1/2)(1 + (1/2)^{1/2})]$. We can then consider the points $\{(\omega_1, 0), (\omega_2, G(\omega_1)), (\omega_3, G(\omega_2)), \ldots\}$, i.e., where each is shifted one place. Taking the convex lower bound of these shifted points gives a convex lower bound for $G_{A|XYE}$ and $G_{AB|X=0,Y=0,E}$. By taking more points in the set $\mathcal{W}$ tighter lower bounds can be obtained.

In the next few sections of this chapter, we shall find ways to compute reliable lower bounds on $G_{A|XYE}(\omega), G_{AB|X=0,Y=0,E}(\omega)$ and $G_{AB|XYE}(\omega)$ for any arbitrary score $\omega \in [\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$.

5.2 H(A|XYE)

Recall from 4.35 that

$$H(A|XYE) \quad = 1 + \textstyle\sum_x p_X(x) H_{\text{bin}}(g(\alpha_x)) - H(E)\,. \tag{5.1}$$

where we define

$$g(\alpha) := \frac{1}{2}\left(1 + \sqrt{2(\lambda_0 - \lambda_3)(\lambda_1 - \lambda_2)\cos(4\alpha) + (\lambda_0 - \lambda_3)^2 + (\lambda_1 - \lambda_2)^2}\right)$$

Using the parameterization (4.28)–(4.31) we have

$$g(\alpha) = \frac{1}{2}\left(1 + R\sqrt{1 + \sin(2\theta)\cos(4\alpha)}\right).$$

We now restrict to the case where $p_X(x) = 1/2$ for $x = 0, 1$. Since $H(E)$ is independent of $\{\alpha_x\}$, we can consider the optimization

$$\begin{aligned} &\min_{\alpha_0,\alpha_1,\beta_0,\beta_1} \; H_{\text{bin}}(g(\alpha_0)) + H_{\text{bin}}(g(\alpha_1)) \\ &\text{subject to} \quad S(\tau_{ABXYE}) = \omega \end{aligned} \tag{5.2}$$

for some fixed values of ω, R and θ.
We proceed to make a series of simplifications of this optimization.

Lemma 22. *The optimization* (5.2) *is equivalent to*

$$\begin{aligned} &\min_{u,v} \; H_{\text{bin}}(g((v+u)/4)) + H_{\text{bin}}(g((v-u)/4)) \\ &\text{s.t.} \quad S(u,v) := \frac{1}{2} + \frac{R}{4}\Bigg(\cos\left(\frac{u}{2}\right)\sqrt{1 + \cos(v)\sin(2\theta)} \\ &\qquad\qquad + \sin\left(\frac{u}{2}\right)\sqrt{1 - \cos(v)\sin(2\theta)}\Bigg) = \omega \\ &\quad 0 \le u \le \pi \end{aligned} \tag{5.3}$$

Proof. Noting that the objective function in (5.2) is independent of Bob's angles (β_0 and β_1), analogously to the derivation of Lemma (18) we can bound the score function

using

$$S(\tau_{ABXYE}) \leq \frac{1}{2} + \frac{R}{2\sqrt{2}}\Big(|\cos(\alpha_0-\alpha_1)|\sqrt{\sin^2(\alpha_0+\alpha_1)\cos^2(\bar{\theta})+\cos^2(\alpha_0+\alpha_1)\sin^2(\bar{\theta})} + |\sin(\alpha_0-\alpha_1)|\sqrt{\cos^2(\alpha_0+\alpha_1)\cos^2(\bar{\theta})+\sin^2(\alpha_0+\alpha_1)\sin^2(\bar{\theta})}\Big).$$

We now substitute $v/2 = \alpha_0+\alpha_1$ and $u/2 = \alpha_0-\alpha_1$ and rearrange (recalling that $\bar{\theta} = \pi/4+\theta$) to give

$$S(\tau_{ABXYE}) \leq \frac{1}{2} + \frac{R}{4}\Big(\Big|\cos\Big(\frac{u}{2}\Big)\Big|\sqrt{1+\cos(v)\sin(2\theta)} + \Big|\sin\Big(\frac{u}{2}\Big)\Big|\sqrt{1-\cos(v)\sin(2\theta)}\Big).$$

Since $G_{A|XYE}(\omega)$ is monotonically increasing in ω (see section 5.5), it follows that we wish to choose the angles to achieve the largest possible score function.
We first note that if $\cos(u/2) < 0$ we can make the substitution $\alpha_0 \mapsto \pi/2+\alpha_1$ and $\alpha_1 \mapsto \alpha_0-\pi/2$ which maintains the objective function, constraint, v and $\sin(u/2)$, while changing the sign of $\cos(u/2)$. In addition, if $\sin(u/2) < 0$, the substitution $\alpha_0 \mapsto \alpha_1$ and $\alpha_1 \mapsto \alpha_0$ maintains the objective function, constraint, v and $\cos(u/2)$ while changing the sign of $\sin(u/2)$. It follows that the maximum of $H_{\text{bin}}(g((v+u)/4))+H_{\text{bin}}(g((v-u)/4))$ for fixed score is obtained when

$$S(\tau_{ABXYE}) = S(u,v) := \frac{1}{2} + \frac{R}{4}\Big(\cos(u/2)\sqrt{1+\cos(v)\sin(2\theta)} + \sin(u/2)\sqrt{1-\cos(v)\sin(2\theta)}\Big)$$

and when both $\cos(u/2) \geq 0$ and $\sin(u/2) \geq 0$, or, alternatively $0 \leq u \leq \pi$. □

Lemma 23. *In the optimization* (5.3) *we can restrict to* $0 \leq u \leq \pi/2$ *and* $0 \leq v \leq \pi/2$ *without affecting the result.*

Proof. Consider a u that satisfies $0 \leq u \leq \pi$. If $\cos(u/2) \geq \sin(u/2)$ then $0 \leq u \leq \pi/2$. Otherwise, consider $u \mapsto \pi-u$, $v \mapsto \pi-v$. This maintains the constraint and the value of the objective function and hence the optimal value can still be obtained, but now with $\cos(u/2) \geq \sin(u/2)$, so we can assume $0 \leq u \leq \pi/2$.
For the restriction on v, first note that transforming $v \mapsto -v$ has no affect on either the objective function or the constraint so we can take $\sin(v) \geq 0$, or $0 \leq v \leq \pi$. If $v > \pi/2$, then $\cos(v+u) < 0$. Furthermore, $\cos(v-u)+\cos(v+u) = 2\cos(v)\cos(u) \leq 0$ and

hence $\cos(v-u) \leq |\cos(v+u)|$. Let $\bar{v} = \pi - v$. We have

$$S(u,v) - S(u,\bar{v}) = \frac{R}{4}\Big((\sin(u/2) - \cos(u/2)) \Big(\sqrt{1-\cos(v)\sin(2\theta)} - \sqrt{1+\cos(v)\sin(2\theta)}\Big)\Big) \leq 0\,,$$

where the inequality follows from $\cos(v) < 0$, $\sin(2\theta) \geq 0$ and $\cos(u/2) \geq \sin(u/2)$. Hence, the mapping $v \mapsto \pi - v$ increases $S(u,v)$.

Consider now the effect on the objective function

$$J(u,v) := \Phi\left(R\sqrt{1+\sin(2\theta)\cos(v+u)}\right) + \Phi\left(R\sqrt{1+\sin(2\theta)\cos(v-u)}\right),$$

where we have used the notation

$$\Phi(x) := H_{\text{bin}}\left(\frac{1}{2} + \frac{x}{2}\right),$$

which will be useful shorthand throughout this thesis. Note that each binary entropy term decreases as its cosine term increases. We have

$$J(u,\bar{v}) := \Phi\left(R\sqrt{1-\sin(2\theta)\cos(v+u)}\right) + \Phi\left(R\sqrt{1-\sin(2\theta)\cos(v-u)}\right).$$

If $\cos(v-u) \leq 0$ then $J(u,\bar{v}) \leq J(u,v)$, so the transformation decreases the objective function.

On the other hand, if $\cos(v-u) \geq 0$ we must have $\cos(v-u) \leq |\cos(v+u)|$ and hence

$$\begin{aligned}\sqrt{1-\sin(2\theta)\cos(v+u)} &\geq \sqrt{1+\sin(2\theta)\cos(v-u)} \\ &\geq \sqrt{1-\sin(2\theta)\cos(v-u)} \\ &\geq \sqrt{1+\sin(2\theta)\cos(v+u)}.\end{aligned}$$

Thus, $J(u,\bar{v}) \leq J(u,v)$ and the transformation decreases the objective function.

Thus, in both cases the transformation decreases the objective function while increasing the score. Using the monotonicity of $G_{A|XYE}(\omega)$ with the score ω (see section 5.5), it follows that we can further reduce the objective function while bringing the score back to its original level. □

Let us turn to the constraint. We have

$$\begin{aligned}\omega &= \frac{1}{2} + \frac{R}{4}\left(\cos(u/2)\sqrt{1+\cos(v)\sin(2\theta)} + \sin(u/2)\sqrt{1-\cos(v)\sin(2\theta)}\right) \\ &= \frac{1}{2} + \frac{R}{2\sqrt{2}}\cos(u/2-\phi)\,,\end{aligned}$$

where $\cos(\phi) = \sqrt{(1+\cos(v)\sin(2\theta))/2}$ and $\sin(\phi) = \sqrt{(1-\cos(v)\sin(2\theta))/2}$. We can rearrange this to $\cos(u/2-\phi) = \sqrt{2}(2\omega-1)/R$ and hence there are two possibilities for u:

$$u_\pm = 2\cos^{-1}\sqrt{(1+\cos(v)\sin(2\theta))/2} \pm 2\cos^{-1}(\sqrt{2}(2\omega-1)/R). \tag{5.4}$$

We can hence remove the constraint and consider the optimizations

$$\min_v \; J(u_\pm(v), v)$$

Summarizing the above analysis we have the following.

Corollary 4. *Let* $\omega \in (3/4, (1+1/\sqrt{2})/2]$ *and*

$$K(R,\theta) = 1 - H(\{\lambda_0(R,\theta), \lambda_1(R,\theta), \lambda_2(R,\theta), \lambda_3(R,\theta)\}),$$

where $\{\lambda_i(R,\theta)\}_{i=0}^3$ *are given by* (4.28)–(4.31) *with* $\delta = \delta^*$. *Defining*

$$\mathcal{D}_\omega = \{(R,\theta,v) : R \in [\sqrt{2}(2\omega-1), 1],\; \theta \in [0, \frac{\pi}{4} - \cos^{-1}\left(1/(\sqrt{2}R)\right)],\; v \in [0, \frac{\pi}{2}]\},$$

we have

$$G_{A|XYE}(\omega) \;=\; \min_{\mathcal{D}_\omega, u\in\{u_+,u_-\}} J(u(v),v)/2 + K(R,\theta)\,. \tag{5.5}$$

5.2.1 Monotonicity properties of the function K

The following monotonicity properties of the function $K(R,\theta)$ will be useful later.

Lemma 24. *For any* $\omega \in (3/4, (1+1/\sqrt{2})/2]$, *and* $(R,\theta) \in \mathcal{D}_\omega$ *we have* $\partial_R K(R,\theta) \geq 0$.

Proof. Note that

$$\frac{\lambda_0\lambda_3}{\lambda_1\lambda_2} = 1$$

and $\lambda_0 > \lambda_3$ and $\lambda_1 > \lambda_2$. We differentiate $K(R,\theta)$ with respect to R

$$\begin{aligned}
\partial_R K(R,\theta) &= \sum_i \left(\log(\lambda_i) + \frac{1}{\ln 2}\right)\frac{\partial\lambda_i}{\partial R} \\
&= \frac{1}{2}\Big((\cos(\theta)+R\cos(2\theta))\,(\log\lambda_0 + \frac{1}{\ln 2}) + (\sin(\theta)-R\cos(2\theta))\,(\log\lambda_1+\frac{1}{\ln 2}) - \\
&\quad (\sin(\theta)+R\cos(2\theta))\,(\log\lambda_2+\frac{1}{\ln 2}) - (\cos(\theta)-R\cos(2\theta))\,(\log\lambda_3+\frac{1}{\ln 2})\Big) \\
&= \frac{1}{2}\left(\log\left(\frac{\lambda_0}{\lambda_3}\right)\cos(\theta) + \log\left(\frac{\lambda_1}{\lambda_2}\right)\sin(\theta) + R\cos(2\theta)\log\left(\frac{\lambda_0\lambda_3}{\lambda_1\lambda_2}\right)\right) \\
&= \log\left(\frac{\lambda_0}{\lambda_3}\right)\cos(\theta) + \log\left(\frac{\lambda_1}{\lambda_2}\right)\sin(\theta) \geq 0
\end{aligned}$$

as claimed. □

We derive a similar result for monotonicity of $K(R,\theta)$ with respect to θ

Lemma 25. *For any $\omega \in (3/4, (1+1/\sqrt{2})/2]$, and $(R,\theta) \in \mathcal{D}_\omega$ we have $\partial_\theta K(R,\theta) \geq 0$.*

Proof. We differentiate $K(R,\theta)$ with respect to θ

$$\begin{aligned}
\partial_\theta K(R,\theta) &= \sum_i \Big(\log(\lambda_i) + \frac{1}{\ln 2} \Big) \frac{\partial \lambda_i}{\partial \theta} \\
&= \sum_i \Big(\log(\lambda_i) \frac{\partial \lambda_i}{\partial \theta} \Big) + \frac{1}{\ln 2} \sum_i \frac{\partial \lambda_i}{\partial \theta} \\
&= \sum_i \Big(\log(\lambda_i) \frac{\partial \lambda_i}{\partial \theta} \Big) \\
&= \frac{1}{2} \left(-\log\left(\frac{\lambda_0}{\lambda_3}\right) R\sin(\theta) + \log\left(\frac{\lambda_1}{\lambda_2}\right) R\cos(\theta) - R^2 \sin(2\theta) \log\left(\frac{\lambda_0 \lambda_3}{\lambda_1 \lambda_2}\right) \right) \\
&= -\frac{R\sin(\theta)}{2} \log\left(\frac{\lambda_0}{\lambda_3}\right) + \frac{R\cos(\theta)}{2} \log\left(\frac{\lambda_1}{\lambda_2}\right)
\end{aligned}$$

We now consider the function

$$f(R,\theta) = \frac{2\ln 2}{R} \partial_\theta K(R,\theta) = -\sin(\theta) \ln\left(\frac{\lambda_0(R,\theta)}{\lambda_3(R,\theta)}\right) + \cos(\theta) \ln\left(\frac{\lambda_0(R,\theta)}{\lambda_3(R,\theta)}\right) \quad (5.6)$$

Note that $\omega \in (3/4, (1+1/\sqrt{2})/2]$ implies $1/\sqrt{2} < R \leq 1$ and $0 \leq \theta \leq \pi/4 - \cos^{-1}(1/(R\sqrt{2}))$, or $0 \leq \theta \leq \pi/4$, $1/\sqrt{2} < R \leq 1/(\cos(\theta)+\sin(\theta))$. We can extend the domain of $f(R,\theta)$ to $0 \leq \theta \leq \pi/4$, $0 \leq R \leq 1/(\cos(\theta)+\sin(\theta))$.
Taking derivative of f with respect to R gives

$$\partial_R f(R,\theta) = \frac{2R^2 \sin(4\theta)}{(1-R^2(\cos(\theta)+\sin(\theta))^2)(1-R^2(\cos(\theta)-\sin(\theta))^2)}. \quad (5.7)$$

Thus, $\partial_R f(R,\theta) > 0$ whenever $\theta \in [0, \frac{\pi}{4}]$. We can then infer that $\partial_\theta K(R,\theta) = \frac{R}{2\ln 2} f(R,\theta) \geq \frac{R}{2\ln 2} f(0,\theta) = 0$. □

5.2.2 LOWER BOUNDING THE OBJECTIVE FUNCTION

In this section, we propose a method to compute the lower bound of the function $G_{.|.E}$ by partitioning the domain. We start by considering an abstract version of the problem, which has the form

$$\min_{\mathbf{x} \in \mathcal{D}} Q(\mathbf{x}) \quad (5.8)$$

where $\mathcal{D} \subset \mathbb{R}^n$ is a compact set and $Q : \mathcal{D} \mapsto \mathbb{R}$ is bounded. Furthermore, we assume we know an upper bound M such that $Q(\mathbf{x}) \leq M$ for all $\mathbf{x} \in \mathcal{D}$.

We use the notation $\mathcal{C}_{\mathbf{a},\mathbf{b}} = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, i.e., $\mathcal{C}_{\mathbf{a},\mathbf{b}}$ is a hyper-cuboid with $\mathbf{a}$ and $\mathbf{b}$ as two opposite vertices. Then let $\mathcal{C} \supseteq \mathcal{D}$ be any hyper-cuboid that completely contains $\mathcal{D}$. We say $\mathcal{P} = \{\mathcal{C}_{\mathbf{a}^i,\mathbf{b}^i}\}_i$ is a partition of $\mathcal{C}$ if

$$\mathcal{C} = \bigcup_i \mathcal{C}_{\mathbf{a}^i,\mathbf{b}^i} \tag{5.9}$$

where $\{\mathcal{C}_{\mathbf{a}^i,\mathbf{b}^i}\}_i$ are cuboids whose intersection has zero volume.
The main idea behind our lower bounds is to find lower bounds on $Q(\mathbf{x})$ that hold on each cuboid and then to take the minimum of all the lower bounds. In some cases these bounds are formed by starting from a corner and using bounds on the derivatives of $Q(\mathbf{x})$ on the cuboid to form a bound that holds across the cuboid. In other cases, we use monotonicity arguments to imply that evaluation at one of the corners lower bounds the whole cuboid. Some of our cuboids lie entirely outside the original domain $\mathcal{D}$. To save calculation we assign the known upper bound on the function as the upper bound on cuboids in our partition that lie outside of $\mathcal{D}$.

5.2.3 Obtaining a lower bound on $G_{A|XYE}$

We now return to our optimization problem (5.5). It is convenient to switch parameterization to use $\eta := \cos^{-1}(\sqrt{2}(2\omega - 1)/R)$ instead of R. Taking $\mathbf{x} = (\eta, \theta, v)$ we rewrite (5.5) as

$$G_{A|XYE}(\omega) = \min_{\mathbf{x}\in\mathcal{D}_\omega, u\in\{u_+,u_-\}} F_1(\eta,\theta,v) + F_2(\eta,\theta,v) + K(R(\eta),\theta) \tag{5.10}$$

where

$$\begin{aligned} F_1(\eta,\theta,v) &= \frac{1}{2} H_{\text{bin}}\left(\frac{1}{2} + \frac{R(\eta)}{2}\sqrt{1 + \cos(u(v) + v)\sin(2\theta)}\right) \\ F_2(\eta,\theta,v) &= \frac{1}{2} H_{\text{bin}}\left(\frac{1}{2} + \frac{R(\eta)}{2}\sqrt{1 + \cos(u(v) - v)\sin(2\theta)}\right) \\ R(\eta) &= \frac{\sqrt{2}(2\omega - 1)}{\cos(\eta)}. \end{aligned}$$

Here the domain $\mathcal{D}_\omega$ is the set

$$\begin{aligned} \mathcal{D}_\omega = \Big\{(\eta,\theta,v) : \eta \in [0, \cos^{-1}\left(\sqrt{2}(2\omega-1)\right)] \\ , \theta \in [0, \frac{\pi}{4} - \cos^{-1}(\cos(\eta)/(4\omega-2))], v \in [0, \frac{\pi}{2}]\Big\}. \end{aligned} \tag{5.11}$$

Define a cuboid $\mathcal{C} \supseteq \mathcal{D}_\omega$ as $[0, \cos^{-1}\left(\sqrt{2}(2\omega-1)\right)] \times [0, \frac{\pi}{4} - \cos^{-1}\left(1/(4\omega-2)\right)] \times [0, \frac{\pi}{2}]$. We then partition $\mathcal{C}$ as follows. We take $\{\eta_i\}_{i=0}^{N+1}$ to be such that $0 = \eta_0 < \eta_1 <$

$\eta_2 \cdots < \eta_{N+1} = \cos^{-1}\left(\sqrt{2}(2\omega-1)\right)$. Similarly define $\{\theta_j^{(i)}\}_{j=0}^{M(i)+1}$ be such that $0 = \theta_0^{(i)} < \theta_1^{(i)} < \cdots < \theta_{M(i)+1}^{(i)} = \frac{\pi}{4} - \cos^{-1}\left(1/(4\omega-2)\right)$ and $\{v_k^{(i,j)}\}_{k=0}^{P(i,j)+1}$ be such that $0 = v_0^{(i,j)} < v_1^{(i,j)} < v_2^{(i,j)} \cdots < v_{P(i,j)+1}^{(i,j)} = \frac{\pi}{2}$. Thus $\mathcal{C} = \bigcup_{i,j,k} \mathcal{C}_{i,j,k}$, where, to streamline the notation, we have used $\mathcal{C}_{i,j,k} := \mathcal{C}_{\mathbf{x}_{i,j,k},\mathbf{x}_{i+1,j+1,k+1}}$ with $\mathbf{x}_{i,j,k} := (\eta_i, \theta_j^{(i)}, v_k^{(i,j)})$. From (5.4) there are two possible functional forms of $u_\pm$. Taking derivatives we find

$$\begin{aligned} \partial_\eta u_\pm &= \pm 2 \\ \partial_\theta u_\pm &= -2\frac{\cos(2\theta)\cos(v)}{\sqrt{1-\sin^2(2\theta)\cos^2(v)}} \in [-2,0] \\ \partial_v u_\pm &= \frac{\sin(2\theta)\sin(v)}{\sqrt{1-\sin^2(2\theta)\cos^2(v)}} \in [0,1]. \end{aligned} \tag{5.12}$$

We return to the problem of deriving an upper bound on the functions F_1 and F_2. To do so, we first need bounds on the functions $\cos(u_\pm \pm v)\sin(2\theta)$. Our bounds use Taylor's theorem, which we first state for convenience.

Theorem 3 (Taylor). *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be compact and $f : \mathcal{D} \to \mathbb{R}$ be differentiable on $\mathcal{D}$, then for all $\mathbf{a}, \mathbf{x} \in \mathcal{D}$ there exists $\mathbf{x}' \in \mathcal{D}$ such that*

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f\Big|_{\mathbf{x}'} \cdot (\mathbf{x} - \mathbf{a}).$$

Thus, we can find lower bounds on f in the domain $\mathcal{D}$ by computing f at any point $\mathbf{a} \in \mathcal{D}$ and the upper-bound $\max_{\mathbf{x}' \in \mathcal{D}} \nabla f(\mathbf{x}')$. We apply this to the functions $\cos(u_\pm(\mathbf{x} \pm v))\sin(2\theta)$ in appendix A.1, where we shall the show all the detailed calculations. For brevity, we write $g_{\pm,y}(\mathbf{x}) = u_\pm(\mathbf{x}) + (-1)^y v$ with $y \in \{0,1\}$. In particular, we show that we have the following bounds for any $\mathbf{x} \in \mathcal{C}_{i,j,k}$:

$$\cos(g_{+,y}(\mathbf{x}))\sin(2\theta) \le \zeta_{+,y}^{i,j,k} \text{ and } \cos(g_{-,y}(\mathbf{x}))\sin(2\theta) \le \zeta_{-,y}^{i,j,k} \tag{5.13}$$

where,

$$\zeta_{+,y}^{i,j,k} := \begin{cases} \left(\cos(g_{+,y}(\mathbf{x}_{i,j,k})) + \Delta_{+,y}\right)\sin\left(2\theta_j^{(i)}\right) & \text{if } \left(\cos(g_{+,y}(\mathbf{x}_{i,j,k})) + \Delta_{+,y}\right) < 0 \\ \left(\cos(g_{+,y}(\mathbf{x}_{i,j,k})) + \Delta_{+,y}\right)\sin\left(2\theta_{j+1}^{(i)}\right) & \text{otherwise} \end{cases}$$

$$\zeta_{-,y}^{i,j,k} := \begin{cases} \left(\cos(g_{-,y}(\mathbf{x}_{i,j,k})) + \Delta_{-,y}\right)\sin\left(2\theta_j^{(i)}\right) & \text{if } \cos(g_{-,y}(\mathbf{x}_{i,j,k})) + \Delta_{-,y} < 0 \\ \left(\cos(g_{-,y}(\mathbf{x}_{i,j,k})) + \Delta_{-,y}\right)\sin\left(2\theta_{j+1}^{(i)}\right) & \text{otherwise} \end{cases}.$$

and

$$\begin{aligned}
\Delta_{+,0} &= \max(2(\theta_{j+1}^{(i)} - \theta_j^{(i)}),\ 2(\eta_{i+1} - \eta_i) + 2(v_{k+1}^{(i,j)} - v_k^{(i,j)})) \\
\Delta_{+,1} &= \max(2(\eta_{i+1} - \eta_i),\ 2(\theta_{j+1}^{(i)} - \theta_j^{(i)}) + (v_{k+1}^{(i,j)} - v_k^{(i,j)}) \\
\Delta_{-,0} &= \max(2(v_{k+1}^{(i,j)} - v_k^{(i,j)}),\ 2(\eta_{i+1} - \eta_i) + 2(\theta_{j+1}^{(i)} - \theta_j^{(i)})) \\
\Delta_{-,1} &= 2(\eta_{i+1} - \eta_i) + 2(\theta_{j+1}^{(i)} - \theta_j^{(i)}) + (v_{k+1}^{(i,j)} - v_k^{(i,j)}).
\end{aligned}$$

With this established we return to the optimization problem (5.10). We define the objective function $Q(\eta, \theta, v) := F_1(\eta, \theta, v) + F_2(\eta, \theta, v) + K(R(\eta), \theta)$, which we want to optimize over $\mathcal{D}_\omega$ and $u \in \{u_+, u_-\}$.

Lemma 26. *Let $\mathcal{P} = \bigcup_{i,j,k} \mathcal{C}_{i,j,k}$ be a partition of $\mathcal{C}$ as specified above. Define $g_{i,j,k}$ and $h_{i,j,k}$ as follows*

$$\begin{aligned}
g_{i,j,k} &:= \frac{1}{2}\Phi\left(R(\eta_{i+1})\sqrt{1+\zeta_{+,0}^{i,j,k}}\right) + \frac{1}{2}\Phi\left(R(\eta_{i+1})\sqrt{1+\zeta_{+,1}^{i,j,k}}\right) + K\left(R(\eta_i), \theta_j^{(i)}\right) \\
h_{i,j,k} &:= \frac{1}{2}\Phi\left(R(\eta_{i+1})\sqrt{1+\zeta_{-,0}^{i,j,k}}\right) + \frac{1}{2}\Phi\left(R(\eta_{i+1})\sqrt{1+\zeta_{-,1}^{i,j,k}}\right) + K\left(R(\eta_i), \theta_j^{(i)}\right).
\end{aligned}$$

Let $M \in \mathbb{R}$ be any upper bound on Q, i.e., $M \geq \max_{\mathbf{x}\in\mathcal{D}} Q(\mathbf{x})$. Then

$$Q(\mathbf{x}) \geq f_{i,j,k} := \begin{cases} \min\{g_{i,j,k}, h_{i,j,k}\} & \text{if } \mathbf{x} \in \mathcal{C}_{i,j,k} \text{ such that } \mathcal{C}_{i,j,k} \cap \mathcal{D} \neq \emptyset \\ M & \text{otherwise.} \end{cases} \tag{5.14}$$

Proof. From Lemmas 24 and 25 we know that $\partial_R K > 0$ and $\partial_\theta K > 0$. In addition, $\partial_\eta K(R(\eta), \theta) = \frac{\sqrt{2}(2\omega-1)\sin(\eta)}{\cos^2(\eta)} \partial_R K(R, \theta)$. Positivity of $\partial_\eta K$ and $\partial_\theta K$, implies $K(R(\eta), \theta) \geq K(R(\eta_i), \theta_j^{(i)})$ within $\mathcal{C}_{i,j,k}$. Furthermore, $H_{\text{bin}}(\frac{1}{2} + \frac{x}{2})$ is decreasing for $x \geq 0$. Since $R(\eta)\sqrt{1+\cos(v \pm u)\sin(2\theta)} > 0$,

$$\begin{aligned}
\Phi\left(R(\eta)\sqrt{1+\cos(u_+ \pm v)\sin(2\theta)}\right) &\geq \Phi\left(R(\eta_{i+1})\sqrt{1+\cos(u_+ \pm v)\sin(2\theta)}\right) \\
&= \Phi\left(R(\eta_{i+1})\sqrt{1+\cos\left(g_{+,(1\mp1)/2}\right)\sin(2\theta)}\right) \\
&\geq \Phi\left(R(\eta_{i+1})\sqrt{1+\zeta_{+,(1\mp1)/2}^{i,j,k}}\right).
\end{aligned}$$

Similarly,

$$\Phi\left(R(\eta_{i+1})\sqrt{1+\cos(u_- \pm v)\sin(2\theta)}\right) \geq \Phi\left(R(\eta_{i+1})\sqrt{1+\zeta_{-,(1\mp1)/2}^{i,j,k}}\right),$$

which establishes the claim. □

Combining the results in this section we obtain the following corollary.

Corollary 5. *Let $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ be fixed. Let $\mathcal{D}_\omega$ be defined as in (5.11) and $\mathcal{P} = \bigcup_{i,j,k} \mathcal{C}_{i,j,k}$ be any partition of the cuboid $\mathcal{C} = [0, \cos^{-1}\left(\sqrt{2}(2\omega - 1)\right)] \times [0, \frac{\pi}{4} - \cos^{-1}\left((4\omega - 2)^{-1}\right)] \times [0, \frac{\pi}{2}]$. Then*

$$G_{A|XYE}(\omega) \geq \min_{i,j,k} f_{i,j,k}. \tag{5.15}$$

where $f_{i,j,k}$ are defined in (5.14).

This means that for fixed ω we can lower bound the randomness by evaluating $f_{i,j,k}$ at all grid points in the relevant cuboid and taking the minimum. This is how our numerical algorithm works (note that the lower bound gets tighter as the number of grid points is increased).

5.2.4 LOWER BOUNDING THE RANDOMNESS RATE

In the previous section, we derived a technique to lower bound the function $G_{A|XYE}(\omega)$ for a fixed value of the score, ω. In Section 4.3, we showed that the asymptotic rate $F_{A|XYE}$ can be computed by taking the convex lower bound on $G_{A|XYE}$. In this section, we construct a lower bound on the function $F_{A|XYE}$ using a lower bound on $G_{A|XYE}$. We start with a general lemma.

Lemma 27. *Let a and b be real numbers, $a < b$ and $\tilde{G} : [a, b] \to \mathbb{R}$ be a lower bound on $G : [a, b] \to \mathbb{R}$. Let $\tilde{F}[a, b] \to \mathbb{R}$ and $F[a, b] \to \mathbb{R}$ be convex lower bounds on $\tilde{G}$ and G respectively. Then $\tilde{F}$ is a lower bound on F.*

Proof. Let M_{ω_0} be the set of probability measures on the interval $[a, b]$ satisfying $\int d\mu(\omega)\omega = \omega_0$.

$$F(\omega_0) = \inf_{\mu \in M_{\omega_0}} \int d\mu(\omega) G(\omega) \tag{5.16}$$

Since $G(\omega) \geq \tilde{G}(\omega)$ for every value of $\omega \in [a, b]$, for every measure $\mu \in M_\omega$ we must have that $\int d\mu(\omega)G(\omega) \geq \int d\mu(\omega)\tilde{G}(\omega)$. Thus

$$F(\omega_0) := \inf_{\mu \in \omega_0} \int d\mu(\omega) G(\omega) \geq \inf_{\mu \in \omega_0} \int d\mu(\omega)\tilde{G}(\omega) \geq \tilde{F}(\omega_0). \qquad \square$$

Since we can only compute our lower bound $G^{\mathcal{P}}_{A|XYE}$ on $G_{A|XYE}$ for a finite set of values of ω, to form a lower bound that holds for all values of ω, we construct a

function $\tilde{G}_{A|XYE}$ as follows. Let $\{\omega_i\}_{i=1}^N$ be an ordered set of values in $[\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ with $\omega_1 = 3/4$ at which we have computed $G^{\mathcal{P}}_{A|XYE}$. We define $\tilde{G}_{A|XYE}(\omega)$ to be equal to $G^{\mathcal{P}}_{A|XYE}(\omega_i)$ for $\omega \in [\omega_i, \omega_{i+1})$, and equal to $G^{\mathcal{P}}_{A|XYE}(\omega_N)$ for $\omega \geq \omega_N$. Because $G_{A|XYE}$ is monotonically increasing in ω (see Lemma 36), it follows that for $\omega \in [\omega_i, \omega_{i+1})$, $G_{A|XYE}(\omega) \geq G_{A|XYE}(\omega_i) \geq G^{\mathcal{P}}_{A|XYE}(\omega_i) = \tilde{G}_{A|XYE}(\omega)$.
A lower bound $\tilde{F}_{A|XYE}$ of $F_{A|XYE}$ can then be formed by taking the convex lower bound of $\tilde{G}^{\mathcal{P}}_{A|XYE}$ (see Lemma 27).

5.3 H(AB|X=0, Y=0, E)

Recall from the section 4.6 that

$$\begin{aligned} H(AB|X=0,Y=0,E) &= 1 + H_{\text{bin}}(2\epsilon_{00}) - H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}) \\ &= H_{\text{bin}}(2\epsilon_{00}) + K(R,\theta). \end{aligned}$$

Thus we need to minimize the above function with respect to the constraint:

$$\sum_{ij} \epsilon_{i,j} = 2(2\omega - 1)\,.$$

5.3.1 Reparameterizing the optimisation problem

We introduce some notation for convenience. Let $\mathbf{x} = (R, \theta, \alpha_0, \alpha_1, \beta_0, \beta_1)$ and define

$$\hat{\epsilon}_{00}(\mathbf{x}) := \cos(\theta)\cos(2\alpha_0 - 2\beta_0) + \sin(\theta)\cos(2\alpha_0 + 2\beta_0) \tag{5.17}$$

$$\hat{\epsilon}_{10}(\mathbf{x}) := \cos(\theta)\cos(2\alpha_1 - 2\beta_0) + \sin(\theta)\cos(2\alpha_1 + 2\beta_0) \tag{5.18}$$

$$\hat{\epsilon}_{01}(\mathbf{x}) := \cos(\theta)\cos(2\alpha_0 - 2\beta_1) + \sin(\theta)\cos(2\alpha_0 + 2\beta_1) \tag{5.19}$$

$$\hat{\epsilon}_{11}(\mathbf{x}) := -\cos(\theta)\cos(2\alpha_1 - 2\beta_1) - \sin(\theta)\cos(2\alpha_1 + 2\beta_1) \tag{5.20}$$

$$K(\mathbf{x}) := K(R,\theta)\,, \tag{5.21}$$

where $K(R,\theta)$ is given in Corollary 4. In this notation, the equation for the constraint is

$$\sum_{ij} \hat{\epsilon}_{i,j} = \frac{4(2\omega - 1)}{R}\,, \tag{5.22}$$

and hence the optimization problem is

$$\begin{aligned} G_{AB|X=0,Y=0,E}(\omega) = \min_{\mathbf{x}\in\mathcal{D}_\omega} \quad & \left(H_{\text{bin}}\left(\frac{1}{2} + \frac{R}{2}\hat{\epsilon}_{00}(\mathbf{x})\right) + K(\mathbf{x})\right) \\ \text{s.t.} \quad & \sum_{ij} \hat{\epsilon}_{ij}(\mathbf{x}) = \frac{4(2\omega-1)}{R}\,, \end{aligned} \tag{5.23}$$

where $\mathcal{D}_\omega = \{R \in [\sqrt{2}(2\omega-1), 1], \theta \in [0, \pi/4 - \cos^{-1}(1/(R\sqrt{2}))], (\alpha_0, \alpha_1, \beta_0, \beta_1) \in \mathbb{R}^4\}$ (see Lemma 18 for the justification of the range of R).
For brevity we use $P(\mathbf{x})$ for the objective function. We call $\mathbf{x} \in \mathcal{D}_\omega$ a solution to the optimization problem (5.23) if $G_{AB|X=0,Y=0,E}(\omega) = P(\mathbf{x})$ and $\mathbf{x}$ satisfies the constraint. For reasons that shall be clear later, we now define the following functions on the extended domain

$$\hat{H}_{\text{bin}}(x) = \begin{cases} H_{\text{bin}}(x) & \text{if } x \in [\frac{1}{2}, 1] \\ 1 & \text{otherwise} \end{cases} \tag{5.24}$$

and

$$\hat{K}(R, \theta) = \begin{cases} K(R, \theta) & \text{if } \sqrt{2}(2\omega - 1) \leq R \leq 1 \text{ and } 0 \leq \theta \leq \frac{\pi}{4} - \cos^{-1}\left(\frac{1}{\sqrt{2}R}\right) \\ 1 & \text{otherwise} \end{cases} \tag{5.25}$$

Here $\hat{K}(R, \theta)$ and $\hat{H}_{\text{bin}}$ both take the value 1 when the functions $K(R, \theta)$ and $H_{\text{bin}}(x)$ are outside the stated range. These values are chosen such that upon extension of the domain, the resulting optimization problem still has the same minimum[1].

Lemma 28. *Let $\mathbf{X}_\omega$ be the set of solutions of (5.23) for some $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$. There exists $\mathbf{x} \in \mathbf{X}_\omega$ such that $\hat{\epsilon}_{00}(\mathbf{x}) > 0$ and $\hat{\epsilon}_{00}(\mathbf{x}) = \max_{i,j} |\epsilon_{ij}(\mathbf{x})|$.*

Proof. We first prove that we can choose

$$|\hat{\epsilon}_{00}(\mathbf{x})| = \max_{i,j} |\hat{\epsilon}_{ij}(\mathbf{x})|. \tag{5.26}$$

From the symmetry of the binary entropy, $H_{\text{bin}}(\frac{1}{2} + \frac{y_1}{2}) < H_{\text{bin}}(\frac{1}{2} + \frac{y_2}{2})$ for $|y_1| > |y_2|$. Now consider the following cases

- Suppose $|\hat{\epsilon}_{00}(\mathbf{x})| < |\hat{\epsilon}_{10}(\mathbf{x})|$: Perform the transformation $\alpha_0 \leftrightarrow \alpha_1$ and $\beta_1 \to \beta_1 + \frac{\pi}{2}$. Under this transformation $\hat{\epsilon}_{00}(\mathbf{x}) \leftrightarrow \hat{\epsilon}_{01}(\mathbf{x})$ and $\hat{\epsilon}_{10}(\mathbf{x}) \leftrightarrow \hat{\epsilon}_{11}(\mathbf{x})$. The CHSH score is hence preserved. This transformation also decreases the objective function, so $\mathbf{x}$ cannot have been an solution to (5.23) prior to the transformation.
- Suppose $|\hat{\epsilon}_{00}(\mathbf{x})| < |\hat{\epsilon}_{01}(\mathbf{x})|$: Perform the transformation $\beta_0 \leftrightarrow \beta_1$ and $\alpha_1 \to \alpha_1 + \frac{\pi}{2}$. Under this transformation $\hat{\epsilon}_{00}(\mathbf{x}) \leftrightarrow \hat{\epsilon}_{10}(\mathbf{x})$ and $\hat{\epsilon}_{01}(\mathbf{x}) \leftrightarrow \hat{\epsilon}_{11}(\mathbf{x})$. Again, this preserves the CHSH score while reducing the objective function.

[1] That $H_{\text{bin}}(x) \leq 1$ and $K(R, \theta) \leq 1$ whenever defined justifies the choice made for defining $\hat{H}_{\text{bin}}(x)$ and $\hat{K}(R, \theta)$.

- Suppose $|\hat{\epsilon}_{00}(\mathbf{x})| < |\hat{\epsilon}_{11}(\mathbf{x})|$: Perform the transformation $\alpha_0 \to \alpha_1 + \frac{\pi}{2}$, $\alpha_1 \to \alpha_0$, $\beta_0 \to \beta_1$ and $\beta_1 \to \beta_0 + \frac{\pi}{2}$. Under this transformation $\hat{\epsilon}_{00}(\mathbf{x}) \leftrightarrow \hat{\epsilon}_{11}(\mathbf{x})$ and $\hat{\epsilon}_{01}(\mathbf{x}) \leftrightarrow \hat{\epsilon}_{10}(\mathbf{x})$. Again, this preserves the CHSH score while reducing the objective function.

Finally, we can show that $\hat{\epsilon}_{00}(\mathbf{x}) > 0$ by observing that for $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ we have

$$R\sum_{i,j}\hat{\epsilon}_{ij}(\mathbf{x}) \quad = \quad 4(2\omega - 1) > 2. \tag{5.27}$$

In addition, for all i, j,

$$R\hat{\epsilon}_{ij}(\mathbf{x}) \quad \leq \quad R(\cos(\theta) + \sin(\theta)) \leq 1, \tag{5.28}$$

where the last inequality follows from (4.28) and (4.29) whose sum can be at most 1. Now suppose that $|\hat{\epsilon}_{00}(\mathbf{x})| = \max_{i,j} |\hat{\epsilon}_{ij}|$ and $\hat{\epsilon}_{00}(\mathbf{x}) < 0$. It follows that

$$\begin{aligned} R\sum_{i,j}\hat{\epsilon}_{ij} &= R\Big(\hat{\epsilon}_{00} + \hat{\epsilon}_{01}\Big) + R\Big(\hat{\epsilon}_{10} + \hat{\epsilon}_{11}\Big) \\ &\leq R\Big(\hat{\epsilon}_{00} + \hat{\epsilon}_{01}\Big) + 2 \\ &\leq 2, \end{aligned}$$

where the first inequality uses (5.28). This is in contradiction with (5.27). □

Lemma 29. *Let $\hat{P}$ be the objective function with extended domain, i.e., $\hat{P}(\mathbf{x}) := \hat{H}_{\text{bin}}\left(\frac{1}{2} + \frac{R\epsilon_{00}(\mathbf{x})}{2}\right) + \hat{K}(\mathbf{x})$, $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$, and let $\mathcal{X}$ be a set such that $\mathcal{D}_\omega \subseteq \mathcal{X} \subseteq \mathbb{R}^6$. Then,*

$$\begin{aligned} G_{AB|X=0,Y=0,E}(\omega) = \min_{\mathbf{x}\in\mathcal{X}} \quad & \hat{P}(\mathbf{x}) \\ \text{s.t.} \quad & \sum_{ij}\hat{\epsilon}_{ij}(\mathbf{x}) = \frac{4(2\omega-1)}{R}, \end{aligned} \tag{5.29}$$

i.e., optimizing over $\hat{P}$ on an extended domain $\mathcal{X}$ gives the same solution as the original optimization (5.23). Furthermore $\exists \mathbf{x} \in \mathcal{D}_\omega$ that is a solution to both optimization problems.

Proof. Let $\mathbf{x}' \in \mathcal{D}_\omega$ achieve the optimal value of P and have $\hat{\epsilon}_{00}(\mathbf{x}') > 0$. [From Lemma 28 such an $\mathbf{x}'$ exists.] Since $\hat{P}(\mathbf{x}) = P(\mathbf{x}) \leq 2$ for all $\mathbf{x} \in \mathcal{D}_\omega$, and $\hat{P}(\mathbf{x}) = 2$ for $\mathbf{x} \in \mathbb{R}^6 \setminus \mathcal{D}_\omega$, $\mathbf{x}'$ must also achieve the optimal value of $\hat{P}$, where it takes the same value. □

5.3.2 SOME SIMPLIFICATIONS

Lemma 30. *Let $\mathbf{X}_\omega$ the set of solutions to the optimization problem (5.23) for some $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$. There exists $\mathbf{x} = (R, \theta, \alpha_0, \alpha_1, \beta_0, \beta_1) \in \mathbf{X}_\omega$ such that the following hold*

- $\sin(\beta_0 + \beta_1) \geq 0$
- $\sin(\beta_0 - \beta_1) \leq 0$

Proof. The expression for the CHSH score satisfies:

$$\sqrt{2}(2\omega - 1) = R\cos(\beta_0 - \beta_1)A_1 + R\sin(\beta_0 - \beta_1)B_1.$$

where,

$$A_1 = \left[\sin(2\alpha_0)\sin(\beta_0 + \beta_1)\cos\left(\frac{\pi}{4} + \theta\right) + \cos(2\alpha_0)\cos(\beta_0 + \beta_1)\sin\left(\frac{\pi}{4} + \theta\right)\right]$$
$$B_1 = \left[\sin(2\alpha_1)\cos(\beta_0 + \beta_1)\cos\left(\frac{\pi}{4} + \theta\right) - \cos(2\alpha_1)\sin(\beta_0 + \beta_1)\sin\left(\frac{\pi}{4} + \theta\right)\right]$$

Let $\alpha_0, \alpha_1, \beta_0, \beta_1$ be optimal parameters. Consider the following algorithm, in which each step is performed in the order shown and depends on the previous ones.

1. If $\sin(\beta_0 + \beta_1) < 0$, then perform the transformations $\beta_i \to -\beta_i$ and $\alpha_i \to -\alpha_i$. We get $\sin(\beta_0 + \beta_1) \geq 0$ from this step onwards.

2. If $\sin(\beta_0 - \beta_1) > 0$ then perform the transformations $\beta_0 \to \beta_0 + \frac{\pi}{2}$, $\beta_1 \to \beta_1 - \frac{\pi}{2}$, $\alpha_i \to \alpha_i + \frac{\pi}{2}$. This step does not affect $\sin(\beta_0 + \beta_1)$. Thus we ensure that $\sin(\beta_0 - \beta_1) \leq 0$ and $\sin(\beta_0 + \beta_1) \geq 0$.

In each step of the algorithm, the values of ϵ_{ij} for all i, j remain the same, hence the CHSH score and the objective function remains invariant throughout. Thus, the transformations maintain optimal parameters. □

5.3.3 Reduction in parameters

To rewrite the optimization in a way that removes the constraint we introduce the following functions

$$\begin{aligned}
\hat{\alpha}_0(\lambda, v, \theta) &:= -2\tan^{-1}\left(\frac{1}{\tan(\lambda)\tan\left(\frac{\pi}{4}+\theta\right)}\right) + \tan^{-1}\left(\frac{1}{\tan(v)\tan\left(\frac{\pi}{4}+\theta\right)}\right) \\
\tilde{\epsilon}(\lambda, v, \theta) &:= \cos(\theta)\cos(\hat{\alpha}_0 - 2v + \lambda) + \sin(\theta)\cos(\hat{\alpha}_0 + 2v - \lambda) \qquad (5.30) \\
z(\lambda, v, \theta) &= \cos(\lambda - v)\left[\sin(\hat{\alpha}_0)\sin(v)\cos\left(\frac{\pi}{4}+\theta\right) + \cos(\hat{\alpha}_0)\cos(v)\sin\left(\frac{\pi}{4}+\theta\right)\right] \\
&\quad + \frac{\sin(\lambda - v)}{\sqrt{2}}\sqrt{1 - \cos(2v)\sin(2\theta)} \\
\hat{R}(\lambda, v, \theta) &:= \frac{\sqrt{2}(2\omega - 1)}{z(\lambda, v, \theta)}. \qquad (5.31)
\end{aligned}$$

We also state the following small lemma for convenience.

Lemma 31. *Let $a, b \in \mathbb{R}$ with $a \neq 0$. The values of $\gamma \in \mathbb{R}$ that form extrema of $a\cos(\gamma) + b\sin(\gamma)$ are*

$$\gamma = \tan^{-1}(b/a) + n\pi \qquad (5.32)$$

for any $n \in \mathbb{Z}$. If $a > 0$ the maxima occur when n is even and the minima when n is odd, and vice-versa if $a < 0$.

Proof. The problem is equivalent to maximizing

$$\frac{a}{\sqrt{a^2+b^2}}\cos(\gamma) + \frac{b}{\sqrt{a^2+b^2}}\sin(\gamma).$$

Let ϕ satisfy $\cos(\phi) = \left(\frac{a}{\sqrt{a^2+b^2}}\right)$ and $\sin(\phi) = \left(\frac{b}{\sqrt{a^2+b^2}}\right)$. Thus, the expression is equivalent to $\cos(\gamma - \phi)$ which has maxima for $\gamma = \phi + 2n\pi$ and minima for $\gamma = \phi + \pi + 2n\pi$ for $n \in \mathbb{Z}$.
If $a > 0$ then this gives maxima for $\gamma = \tan^{-1}(b/a) + 2n\pi$ and minima for $\gamma = \tan^{-1}(b/a) + (2n+1)\pi$.
Alternatively, if $a < 0$ then this gives maxima for $\gamma = \tan^{-1}(b/a) + (2n+1)\pi$ and minima for $\gamma = \tan^{-1}(b/a) + 2n\pi$. □

Lemma 32. *Let $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ and*

$$\mathcal{D}'_\omega = \left\{(\lambda, v, \theta) \in \mathbb{R}^3 : \lambda \in [0, \pi], v \in [0, \pi], \theta \in [0, \frac{\pi}{4} - \cos^{-1}\left(1/(4\omega - 2)\right)]\right\},$$

then

$$G_{AB|X=0,Y=0,E}(\omega) = \inf_{\mathcal{D}'_\omega} \left(\hat{H}_{\text{bin}} \left(\frac{1}{2} + \frac{\hat{R}(\lambda, v, \theta)\tilde{\epsilon}(\lambda, v, \theta)}{2} \right) + \hat{K}(\hat{R}, \theta) \right) \tag{5.33}$$

Proof. Start from the form of G in Lemma 29. The objective function $\hat{P}$ is independent of the parameters α_1 and β_1, and, as shown in Lemma 29, the optimum is achieved for some $\mathbf{x} \in \mathcal{D}_\omega$. Because the function $G_{AB|X=0,Y=0,E}(\omega)$ is increasing in ω (see Lemma 37), the optimal values of the parameters α_1 and β_1 must maximize the CHSH score. Recall that the score can be related to α_i and β_i by

$$\sqrt{2}(2\omega - 1) = R\cos(\beta_0 - \beta_1)A_1 + R\sin(\beta_0 - \beta_1), \tag{5.34}$$

where

$$A_1 = \left[\sin(2\alpha_0)\sin(\beta_0 + \beta_1)\cos\left(\frac{\pi}{4} + \theta\right) + \cos(2\alpha_0)\cos(\beta_0 + \beta_0)\sin\left(\frac{\pi}{4} + \theta\right) \right]$$
$$B_1 = \left[\sin(2\alpha_1)\cos(\beta_0 + \beta_1)\cos\left(\frac{\pi}{4} + \theta\right) - \cos(2\alpha_1)\sin(\beta_0 + \beta_1)\sin\left(\frac{\pi}{4} + \theta\right) \right].$$

Consider maximizing this over α_1. From Lemma 30 we can assume $\sin(\beta_0 - \beta_1) \leq 0$, so we want to minimize the second term in square brackets in (5.34). This has the form of the expression in Lemma 31. Since the sine and cosine of $\pi/4 + \theta$ are both positive, and from Lemma 30 we can assume $\sin(\beta_0 + \beta_1) \geq 0$, the minima of the square bracket (and hence maxima overall) occur for

$$2\alpha_1 = -\tan^{-1}\left(\cot(\beta_0 + \beta_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) + 2n\pi. \tag{5.35}$$

The CHSH score is symmetric in the parameters for Alice and Bob, so we can re-write it as

$$\sqrt{2}(2\omega - 1) = R\cos(\alpha_0 - \alpha_1)\hat{A}_1 + R\sin(\alpha_0 - \alpha_1)\hat{B}_1$$

where,

$$\hat{A}_1 = \left[\sin(2\beta_0)\sin(\alpha_0 + \alpha_1)\cos\left(\frac{\pi}{4} + \theta\right) + \cos(2\beta_0)\cos(\alpha_0 + \alpha_0)\sin\left(\frac{\pi}{4} + \theta\right) \right]$$
$$\hat{B}_1 = \left[\sin(2\beta_1)\cos(\alpha_0 + \alpha_1)\cos\left(\frac{\pi}{4} + \theta\right) - \cos(2\beta_1)\sin(\alpha_0 + \alpha_1)\sin\left(\frac{\pi}{4} + \theta\right) \right].$$

If we now maximize over β_1, from Lemma 30 the solutions either satisfy

$$2\beta_1 = -\tan^{-1}\left(\cot(\alpha_0 + \alpha_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) + 2n\pi \quad \text{or}$$
$$2\beta_1 = -\tan^{-1}\left(\cot(\alpha_0 + \alpha_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) + (2n+1)\pi$$

for $n \in \mathbb{Z}$. (Which one holds depends on the signs of $\sin(\alpha_0 - \alpha_1)$ and $\sin(\alpha_0 + \alpha_1)$.) In both cases, $\tan(2\beta_1) = \cot(\alpha_0 + \alpha_1)\cot\left(\frac{\pi}{4} + \theta\right)$.
By symmetry (and because we can take $\sin(\alpha_0 - \alpha_1) \leq 0$ and $\sin(\alpha_0 + \alpha_1) \geq 0$ from Lemma 30) the maxima of this over β_1 occur for

$$2\beta_1 = -\tan^{-1}\left(\cot(\alpha_0 + \alpha_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) + 2n\pi. \tag{5.36}$$

Rearranging gives

$$\tan(\alpha_0 + \alpha_1) = -\cot(2\beta_1)\cot\left(\frac{\pi}{4} + \theta\right), \tag{5.37}$$

and hence

$$\alpha_0 \;=\; -\alpha_1 - \tan^{-1}\left(\cot(2\beta_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) + n\pi$$

for $n \in \mathbb{Z}$. Using (5.35) we find

$$2\alpha_0 = \tan^{-1}\left(\cot(\beta_0 + \beta_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) - 2\tan^{-1}\left(\cot(2\beta_1)\cot\left(\frac{\pi}{4} + \theta\right)\right) + 2n\pi. \tag{5.38}$$

The proof proceeds as follows. We use (5.35) to eliminate α_1 from the constraint, noting that the value of n in (5.35) does not change the value so we can take $n = 0$. We then use (5.38) to reparameterize the objective function in terms of β_1 instead of α_0 (again the value of n in (5.38) makes no difference and we take $n = 0$). The parameters that remain are hence β_0, β_1, R and θ. We then reparameterize using $v = \beta_0 + \beta_1$ and $\lambda = 2\beta_1$, so that both the constraint and objective function are written in terms of v, λ, R and θ. We then use the constraint to write R in terms of the other parameters, reducing the objective function to an unconstrained optimization over v, λ and θ.
At this stage v and λ range over all reals, which can readily be restricted to $[0, 2\pi]$. In fact, we can restrict both to $[0, \pi]$ by noting that Lemma 30 shows that it suffices to take $\sin(v) = \sin(\beta_0 + \beta_1) \geq 0$ hence $v \in [0, \pi]$. We then consider the transformation $(\lambda \mapsto 2\pi - \lambda, v \mapsto \pi - v)$. We find

$$\begin{aligned}
\hat{\alpha}_0(2\pi - \lambda, \pi - v, \theta) &= -\hat{\alpha}_0(\lambda, v, \theta) \\
\tilde{\epsilon}(2\pi - \lambda, \pi - v, \theta) &= \tilde{\epsilon}(\lambda, v, \theta) \\
\hat{R}(2\pi - \lambda, \pi - v, \theta) &= \hat{R}(\lambda, v, \theta)\,,
\end{aligned}$$

from which it follows that we can restrict both λ and v to the range $[0, \pi]$. Finally, the original range of θ is $[0, \pi/4 - \cos^{-1}(1/(R\sqrt{2})]$, with $R \in [\sqrt{2}(2\omega - 1), 1]$, hence the largest θ that needs to be considered for a given ω is $\pi/4 - \cos^{-1}(1/(4\omega - 2))$. Since we

are using the functions with extended domain, it does not matter that we allow the range of θ to potentially be incompatible with the value of $\hat{R}$. This gives the optimization claimed in (5.33). □

5.3.4 LOWER BOUNDING THE FUNCTION

Consider a partition $\mathcal{P}$ of $\mathcal{D}'_\omega$. Let $\mathcal{C}_{i,j,k}$ be a cuboid (with i label corresponding to λ, j label for v and k label for θ). Again using Taylor's theorem, we bound the objective function

$$\left(\hat{H}_{\text{bin}}\left(\frac{1}{2}+\frac{\hat{R}(\lambda,v,\theta)\tilde{\epsilon}(\lambda,v,\theta)}{2}\right)+\hat{K}(\hat{R},\theta)\right)$$

in the cuboid $\mathcal{C}_{i,j,k}$, by upper-bounding the absolute values of the derivatives in the cuboid. Let us define $z(\lambda,v,\theta)$ to be the denominator in (5.31), i.e.,

$$\begin{aligned} z(\lambda,v,\theta) \;:=\; \cos(v-\lambda)\Big[\sin(\hat{\alpha}_0)\sin(v)\cos\Big(\frac{\pi}{4}+\theta\Big)+\cos(\hat{\alpha}_0)\cos(v)\sin\Big(\frac{\pi}{4}+\theta\Big)\Big] \\ -\quad \frac{\sin(v-\lambda)}{\sqrt{2}}\sqrt{1-\cos(2v)\sin(2\theta)}. \end{aligned}$$

In appendix A.2 we use taylor's theorem and some monotonicity results to find the parameters z_λ, z_v, z_θ and $\epsilon_\lambda, \epsilon_v, \epsilon_\theta$ defined as the following upper-bounds :

$$\begin{aligned} z_\lambda &\geq \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} |\partial_\lambda z|, \quad z_v \geq \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} |\partial_v z| \quad \text{and} \quad z_\theta \geq \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} |\partial_\theta z| \\ \epsilon_\lambda &\geq \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} |\partial_\lambda \tilde{\epsilon}|, \quad \epsilon_v \geq \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} |\partial_v \tilde{\epsilon}| \quad \text{and} \quad \epsilon_\theta \geq \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} |\partial_\theta \tilde{\epsilon}|. \end{aligned}$$

Let $\Delta z = z_\lambda(\lambda_{i+1}-\lambda_i) + z_v(v^{(i)}_{j+1}-v^{(i)}_j) + z_\theta(\theta^{(i,j)}_{k+1}-\theta^{(i,j)}_k)$, then in $\mathcal{C}_{i,j,k}$.

$$R^{i,j,k}_{\min} \;:=\; \frac{\sqrt{2}(2\omega-1)}{z(\lambda_i,v^{(i)}_j,\theta^{(i,j)}_k)+\Delta z} \leq \hat{R}(\lambda,v,\theta) = \frac{\sqrt{2}(2\omega-1)}{z(\lambda_i,v^{(i)}_j,\theta^{(i,j)}_k)} \tag{5.39}$$

$$R^{i,j,k}_{\max} \;:=\; \frac{\sqrt{2}(2\omega-1)}{z(\lambda_i,v^{(i)}_j,\theta^{(i,j)}_k)-\Delta z} \geq \hat{R}(\lambda,v,\theta) \tag{5.40}$$

Also let $\Delta\epsilon := \epsilon_\lambda(\lambda_{i+1}-\lambda_i) + \epsilon_v(v^{(i)}_{j+1}-v^{(i)}_j) + \epsilon_\theta(\theta^{(i,j)}_{k+1}-\theta^{(i,j)}_k)$, then in $\mathcal{C}_{i,j,k}$ we have

$$\tilde{\epsilon}(\lambda,v,\theta) \leq \epsilon^{i,j,k}_{\max} := \tilde{\epsilon}(\lambda_i,v_j,\theta_k)+\Delta\epsilon \tag{5.41}$$

For each cuboid we define a continuous function $g_{i,j,k} : \mathcal{C}_{i,j,k} \to \mathbb{R}$ such that $g_{i,j,k}(\mathbf{x}) \leq \hat{P}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{C}_{i,j,k}$. Then we lower bound $G_{AB|X=0,Y=0,E}$ by using the following.

Lemma 33. *Let*

$$g_{i,j,k} := \hat{H}_{\text{bin}}\left(\frac{1}{2} + \frac{R^{i,j,k}_{\max}\epsilon^{i,j,k}_{\max}}{2}\right) + \hat{K}(R^{i,j,k}_{\min}, \theta^{(i,j)}_k). \tag{5.42}$$

Then $\hat{P}(\mathbf{x}) \geq g_{i,j,k}$ *for all* $\mathbf{x} \in \mathcal{C}_{i,j,k}$.

Proof. By definition, we have $R^{i,j,k}_{\max}\epsilon^{i,j,k}_{\max} \geq \hat{R}(\lambda, v, \theta)\tilde{\epsilon}(\lambda, v, \theta)$ for all $\mathbf{x} \in \mathcal{C}_{i,j,k}$. Using the monotonicity of the function $\hat{H}_{\text{bin}}(\frac{1}{2} + \frac{x}{2})$, we obtain $\hat{H}_{\text{bin}}(\frac{1}{2} + \frac{R^{i,j,k}_{\max}\epsilon^{i,j,k}_{\max}}{2}) \leq \hat{H}_{\text{bin}}(\frac{1}{2} + \frac{\hat{R}(\lambda,v,\theta)\tilde{\epsilon}(\lambda,v,\theta)}{2})$. Similarly, the monotonicity of $\hat{K}(R, \theta)$ with respect to R and θ (see Lemmas 24 and 25) implies $\hat{K}(R(\lambda, v, \theta), \theta) \geq \hat{K}(R_{\min}, \theta^{(i,j)}_k)$ for all $\mathbf{x} \in \mathcal{C}_{i,j,k}$. These imply the claim. □

Combining all the results in this section, we have the following

Corollary 6. *Let* $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ *be fixed. Let* $\mathcal{D}'_\omega = \{(\lambda, v, \theta) \in \mathbb{R}^3 : \lambda \in [0, \pi], v \in [0, \pi], \theta \in [0, \frac{\pi}{4} - \cos^{-1}\left(\frac{1}{2(2\omega-1)}\right)]\}$ *and* $\mathcal{P} = \bigcup_{i,j,k} \mathcal{C}_{i,j,k}$ *be a partition of any cuboid* $\mathcal{C} \supseteq \mathcal{D}'(\omega)$ *as specified above. Then*

$$G_{AB|X=0,Y=0,E}(\omega) \geq \min_{i,j,k} g_{i,j,k} \tag{5.43}$$

where $g_{i,j,k}$ *are defined in* (5.42).

Proof. This is a direct consequence of Lemmas 29 and 33. □

5.4 H(AB|XYE)

From section 4.7, recall that

$$H(AB|XYE) = 1 + \sum_{xy} p_{XY}(x, y) H_{\text{bin}}(2\epsilon_{xy}) - H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}). \tag{5.44}$$

So, we set $\delta = \delta^*$, as argued in the section 4.7, and arrive at the optimization problem:

$$\begin{aligned} G_{AB|XYE}(\omega) = \min \quad & \sum_{ij} \frac{1}{4} H_{\text{bin}}(\epsilon_{ij}) + K(R, \theta) \\ \text{s.t.} \quad & \sum_{i,j} \epsilon_{ij} = 2(2\omega - 1) \\ & R\cos(\theta) + R\sin(\theta) \leq 1 \end{aligned}$$

5.4.1 OPTIMIZATION OF H(AB|XYE)

We introduce some notation for convenience. Let $\mathbf{x} = (R, \theta, \alpha_0, \alpha_1, \beta_0, \beta_1)$ and define

$$\begin{aligned}
\hat{\epsilon}_{00}(\mathbf{x}) &:= R\cos(\theta)\cos(2\alpha_0 - 2\beta_0) + R\sin(\theta)\cos(2\alpha_0 + 2\beta_0) \\
\hat{\epsilon}_{10}(\mathbf{x}) &:= R\cos(\theta)\cos(2\alpha_1 - 2\beta_0) + R\sin(\theta)\cos(2\alpha_1 + 2\beta_0) \\
\hat{\epsilon}_{01}(\mathbf{x}) &:= R\cos(\theta)\cos(2\alpha_0 - 2\beta_1) + R\sin(\theta)\cos(2\alpha_0 + 2\beta_1) \\
\hat{\epsilon}_{11}(\mathbf{x}) &:= -R\cos(\theta)\cos(2\alpha_1 - 2\beta_1) - R\sin(\theta)\cos(2\alpha_1 + 2\beta_1) \\
K(\mathbf{x}) &:= K(R, \theta)\,,
\end{aligned}$$

where $K(R, \theta)$ is given in Corollary 4. In this notation, the equation for the constraint is

$$\sum_{ij} \hat{\epsilon}_{i,j} \;=\; 4(2\omega - 1)\,. \tag{5.45}$$

Hence we obtain the optimization problem:

$$\begin{aligned}
G_{AB|XYE}(\omega) = \min_{\mathbf{x}\in\mathcal{D}_\omega} \quad & \sum_{ij} \frac{1}{4}\Phi\left(\hat{\epsilon}_{ij}(\mathbf{x})\right) + K(\mathbf{x}) \\
\text{s.t.} \quad & \sum_{ij} \hat{\epsilon}_{ij}(\mathbf{x}) = 4(2\omega - 1)\,,
\end{aligned} \tag{5.46}$$

where,

$$\mathcal{D}_\omega = \{R \in [\sqrt{2}(2\omega - 1), 1], \theta \in [0, \pi/4 - \cos^{-1}(1/(R\sqrt{2}))], (\alpha_0, \alpha_1, \beta_0, \beta_1) \in \mathbb{R}^4\}$$

(see Lemma 18 for the justification of the range of R). Due to the monotonocity of the function $G_{AB|XYE}(\omega)$ (see lemma 38).

In the following section, we will demonstrate how to obtain reliable lower bounds for $G_{AB|XYE}$ (converging in the asymptotic limit) by considering a sequence of polynomial optimization problems. To initiate this process of finding appropriate lower bounds, we first introduce the following modified optimization problem:

$$\begin{aligned}
G^{(n)}_{AB|XYE}(\omega) = \min_{\mathbf{x}\in\mathcal{D}_\omega} \quad & \sum_{ij} \frac{1}{4}\Phi_n\left(R\hat{\epsilon}_{ij}(\mathbf{x})\right) + K(\mathbf{x}) \\
\text{s.t.} \quad & \sum_{ij} \hat{\epsilon}_{ij}(\mathbf{x}) \geq 4(2\omega - 1).
\end{aligned} \tag{5.47}$$

Here, $\Phi_n(x)$ is a polynomial that serves as a lower bound to the function $\Phi(x)$ in the range $[-1, 1]$. Conveniently, we have a set of polynomial lower bounds for the function $\Phi(x)$, discussed in detail during the discussion of rates for semi Device Independent

protocols in Section 8.9. These functions possess an appealing property, in that they converge to $\Phi(x)$ from below. Specifically, for all $x \in [0,1]$, the following holds:

$$\lim_{n\to\infty} (\Phi(x) - \Phi_n(x)) = 0.$$

If we chose $\Phi_n(x)$ functions to be these special sequence of functions, then we obtain a sequence $\{G^{(n)}_{AB|XYE}\}_n$ that must converge to the function $G_{AB|XYE}$ from below.

5.4.2 Partitioning the domain

Let $\mathcal{P}$ be a partition of the set

$$\mathcal{D}'_\omega := \{(R,\theta) \in \mathbb{R}^2 : \quad R \in [\sqrt{2}(2\omega-1), 1], \theta \in [0, \pi/4 - \cos^{-1}(1/(R\sqrt{2}))]\}.$$

Let $\mathcal{C}_{a,b}$ be a cuboid $[R_a, R_{a+1}] \times [\theta_b^{(a)}, \theta_{b+1}^{(a)}]$ and $\mathbf{x}_{a,b} := (R_a, \theta_b^{(a)})$. Let $\Delta R := R_{a+1} - R_a$ and $\Delta\Theta := \theta_{b+1}^{(a)} - \theta_b^{(a)}$.
Furthermore define

$$\begin{aligned}
\hat{\epsilon}^{(a,b)}_{00}(\mathbf{x}) &:= R_a \cos\left(\theta_b^{(a)}\right)\cos(2\alpha_0 - 2\beta_0) + R_a \sin\left(\theta_b^{(a)}\right)\cos(2\alpha_0 + 2\beta_0)) \\
\hat{\epsilon}^{(a,b)}_{10}(\mathbf{x}) &:= R_a \cos\left(\theta_b^{(a)}\right)\cos(2\alpha_1 - 2\beta_0) + R_a \sin\left(\theta_b^{(a)}\right)\cos(2\alpha_1 + 2\beta_0) \\
\hat{\epsilon}^{(a,b)}_{01}(\mathbf{x}) &:= R_a \cos\left(\theta_b^{(a)}\right)\cos(2\alpha_0 - 2\beta_1) + R_a \sin\left(\theta_b^{(a)}\right)\cos(2\alpha_0 + 2\beta_1) \\
\hat{\epsilon}^{(a,b)}_{11}(\mathbf{x}) &:= -R_a \cos\left(\theta_b^{(a)}\right)\cos(2\alpha_1 - 2\beta_1) - R_a \sin\left(\theta_b^{(a)}\right)\cos(2\alpha_1 + 2\beta_1)
\end{aligned}$$

On a cuboid $\mathcal{C}_{a,b}$ define the restriction of the optimization problem (5.47) on the set $\mathcal{C}_{a,b}$ as:

$$\begin{aligned}
G^{(n)}_{\mathcal{C}_{a,b}}(\omega) := \min_{(R,\theta)\in\mathcal{C}_{a,b}} & \sum_{i,j} \Phi_n\left(\hat{\epsilon}_{i,j}\right) + K(R,\theta) \\
\text{s.t.} \quad & \sum_{ij} \hat{\epsilon}_{ij}(\mathbf{x}) \geq 4(2\omega - 1) \\
& \forall i,j : \alpha_i, \beta_j \in \mathbb{R}.
\end{aligned}$$

We now seek a lower bound $g^{(n)}_{a,b}$ for the function $G^{(n)}_{\mathcal{C}_{a,b}}$, which will allow us to compute a lower bound on the function $G^{(n)}_{AB|XYE}$ using Lemma 29. The following result finds appropriate lower bounds $g^{(n)}_{a,b}$.

Lemma 34. *Let $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ be fixed, and $\mathcal{P} = \bigcup_{a,b} \mathcal{C}_{a,b}$ be a partition of any cuboid $\mathcal{C} \supseteq \mathcal{D}'(\omega)$ as specified above. For any $\mathcal{C}_{a,b}$ in the partition, we define:*

$$\begin{aligned} g_{a,b} := \min_{(R,\theta)\in\mathcal{C}_{a,b}} & \frac{1}{4}\sum_{i,j}\Phi_n\left(\hat{\epsilon}_{i,j}^{(a,b)}\right) + K(R_i,\theta_j^{(i)}) - 2|\Phi_n'(1)|\sqrt{\Delta R^2 + \Delta\theta^2} \\ s.t. \quad & \sum_{ij}\hat{\epsilon}_{ij}(\mathbf{x}) \geq 4(2\omega-1) - 2\sqrt{(\Delta R)^2 + (\Delta\theta)^2} \\ & \forall i,j : \alpha_i, \beta_j \in \mathbb{R}. \end{aligned}$$

Then, for every $\mathcal{C}_{a,b}$ in the partition, we have $g_{a,b} \leq G^{(n)}_{\mathcal{C}_{a,b}}$.

Proof. We invoke lemma 58 (see appendix A.3) to compute the lower bound on the optimization problem for (5.48).
Let $\mathbf{x} = (R, \theta)$ and $\mathbf{y} = (\alpha_0, \alpha_1, \beta_0, \beta_1)$ be vectors corresponding to the parameters in our domain. This allows us to compare optimization problem (5.48) to the optimization problem in lemma 58 by identifying functions $g(\mathbf{x}, \mathbf{y}) = \sum_{a,b}\Phi_n(\hat{\epsilon}_{i,j})$ and $h(\mathbf{x}) = K(R,\theta)$. The constraint function $f_1(\mathbf{x},\mathbf{y}) = \sum_{i,j}\hat{\epsilon}_{i,j}(\mathbf{x},\mathbf{y})$.
The monotonicity of $\hat{K}(R,\theta)$ with respect to R and θ (see Lemmas 24 and 25) implies $\hat{K}(R,\theta) \geq \hat{K}(R_a, \theta_b^{(a)})$ for all $\mathbf{x} \in \mathcal{C}_{a,b}$. This inspires us to choose $\mathbf{x}_0 = (R_a, \theta_b^{(a)})$.
Now, we find the $\Delta_{\max}, g_{\max}$ and $f_{\max}$. To find $\Delta_{\max}$ note that every $\mathbf{x} \in \mathcal{C}_{a,b}$, the following holds:

$$||\Delta_{\mathbf{x}_0}(\mathbf{x})|| = \sqrt{(R-R_a)^2 + (\theta - \theta_b^{(a)})^2} \leq \sqrt{(\Delta R)^2 + (\Delta\theta)^2}.$$

Now, we can bound the gradients:

$$\begin{aligned} |\partial_R\hat{\epsilon}_{i,j}| &\leq (\cos(\theta) + \sin(\theta)) \leq \sqrt{2} \\ |\partial_\theta\hat{\epsilon}_{i,j}| &\leq R(\cos(\theta) + \sin(\theta)) \leq \sqrt{2} \\ |\partial_R g| &\leq |\Phi_n'(1)||\partial_R\hat{\epsilon}_{i,j}| \leq \sqrt{2}|\Phi_n'(1)| \\ |\partial_\theta g| &\leq |\Phi_n'(1)||\partial_\theta\hat{\epsilon}_{i,j}| \leq \sqrt{2}|\Phi_n'(1)|, \end{aligned}$$

where we have used the fact that the functions $\max_{x\in[-1,1]} : |\Phi_n'(x)| = |\Phi_n'(1)| = |\Phi_n'(-1)|$. Combining the values of the derivatives gives $g_{\max} = 2\Phi'(1)$ and $f_{\max} = 2$. □

Combining all the results in this section, we have the following

Corollary 7. *Let $\omega \in (\frac{3}{4}, \frac{1}{2} + \frac{1}{2\sqrt{2}}]$ be fixed. Let $\mathcal{D}'_\omega := \{(R,\theta) \in \mathbb{R}^2 : R \in [\sqrt{2}(2\omega-1), 1], \theta \in [0, \pi/4 - \cos^{-1}(1/(R\sqrt{2}))]\}$ and $\mathcal{P} = \bigcup_{a,b}\mathcal{C}_{a,b}$ be a partition of*

any cuboid $\mathcal{C} \supseteq \mathcal{D}'(\omega)$ as specified above. Then

$$G_{AB|XYE}(\omega) \geq \min_{a,b} g_{a,b}, \tag{5.48}$$

where $g_{a,b}$ are defined in Lemma 34.

Proof. This is a direct consequence of Lemmas 29 and 34. □

The primary motivation for going through the various steps and simplifications is to recognize that $g_{a,b}^{(n)}$ can be easily cast to a polynomial optimization problem. This realization stems from the fact that both, the objective function and the constraints involve the sine and cosine of variables $\alpha_0, \alpha_1, \beta_0, \beta_1 \in \mathbb{R}$. Therefore, we introduce new variables $\cos(\alpha_i) = x_{\alpha_i}$ and $\sin(\alpha_i) = y_{\alpha_i}$, along with analogous variables for β_j. By doing so, both the objective function and the constraints in the optimization problem 5.48 can be expressed in terms of polynomials involving $x_{\alpha_i}, y_{\alpha_i}, x_{\beta_j}, y_{\beta_j}$. Further, we should introduce the additional constraints $x_{\alpha_i}^2 + y_{\alpha_i}^2 = 1$ and $x_{\beta_j}^2 + y_{\beta_j}^2 = 1$ to ensure that x and y correspond to cosine and sine of a valid angle. Consequently, we can determine reliable lower bounds for $g_{a,b}$ using the SDP based techniques to solve polynomial optimization problems as discussed in the section 2.3. These lower bounds converge asymptotically to the actual value of $g_{a,b}$. Furthermore, refining the partition leads to a reduction in the size of cuboid dimensions ΔR and $\Delta\theta$, giving arbitrary tight lower bounds on $G_{AB|XYE}$.

5.5 MONOTONICITY OF RATES

In this section we prove the monotonicity of the functions $G_{A|XYE}(\omega)$, $G_{AB|00E}(\omega)$ and $G_{AB|XYE}(\omega)$. There is a common part to the proofs, which we first establish.

Lemma 35. *Let $\lambda_0(R,\theta), \lambda_1(R,\theta), \lambda_2(R,\theta)$ and $\lambda_3(R,\theta)$ be the eigenvalues of a Bell-diagonal state $\rho_{A'B'}$ as in (4.28)–(4.31) in the case where $\delta = \frac{R^2}{4}\cos(2\theta)$. Then*

$$\frac{\partial}{\partial R}\left(H_{\text{bin}}(\lambda_0+\lambda_1) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\})\right) > 0. \tag{5.49}$$

Proof.

$$\frac{\partial}{\partial R}\Big(H_{\text{bin}}(\lambda_0+\lambda_1)\Big) \;=\; -\log(\lambda_0+\lambda_1)\frac{\partial}{\partial R}(\lambda_0+\lambda_1) - \log(\lambda_2+\lambda_3)\frac{\partial}{\partial R}(\lambda_2+\lambda_3)$$

The equality above follows from the fact that $1-\lambda_1-\lambda_0 = \lambda_2+\lambda_3$ and thus $H_{\text{bin}}(\lambda_0+\lambda_1) = -(\lambda_0+\lambda_1)\log(\lambda_1+\lambda_0) - (\lambda_2+\lambda_3)\log(\lambda_2+\lambda_3)$. We also have that

$$\frac{\partial}{\partial R}H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\}) = -\sum_i \log\lambda_i \frac{\partial\lambda_i}{\partial R}. \tag{5.50}$$

For convenience, we write

$$G = H_{\text{bin}}(\lambda_0 + \lambda_1) - H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}).$$

Adding the derivatives, we define

$$\begin{aligned}
\frac{\partial}{\partial R}\Big(G\Big) &= \log\left(\frac{\lambda_0}{\lambda_0+\lambda_1}\right)\frac{\partial \lambda_0}{\partial R} + \log\left(\frac{\lambda_1}{\lambda_0+\lambda_1}\right)\frac{\partial \lambda_1}{\partial R} \\
&\quad + \log\left(\frac{\lambda_2}{\lambda_3+\lambda_2}\right)\frac{\partial \lambda_2}{\partial R} + \log\left(\frac{\lambda_3}{\lambda_2+\lambda_3}\right)\frac{\partial \lambda_3}{\partial R} \\
&= \log_2\left(\frac{\lambda_0}{\lambda_0+\lambda_1}\right)\frac{\partial}{\partial R}(\lambda_0+\lambda_2) + \log_2\left(\frac{\lambda_1}{\lambda_0+\lambda_1}\right)\frac{\partial}{\partial R}(\lambda_1+\lambda_3) \\
&= \log\left(\frac{\lambda_0}{\lambda_1}\right)\frac{\partial}{\partial R}(\lambda_0+\lambda_2) = \log\left(\frac{\lambda_0}{\lambda_1}\right)\frac{\cos(\theta)-\sin(\theta)}{2} \\
&\geq 0, \qquad (5.51)
\end{aligned}$$

where the second equality follows from the fact that for Bell-diagonal states parameterized by $\delta = \frac{R^2}{4}\cos(2\theta)$, the eigenvalues obey

$$\frac{\lambda_0}{\lambda_0+\lambda_1} = \frac{\lambda_2}{\lambda_2+\lambda_3} \quad \text{and} \quad \frac{\lambda_1}{\lambda_0+\lambda_1} = \frac{\lambda_3}{\lambda_2+\lambda_3}$$

and the inequality comes from the parameterization. □

Lemma 36. *For $\omega \in (\frac{3}{4}, \frac{1}{2}(1+\frac{1}{\sqrt{2}}))$ and any distribution p_{XY}, the function $G_{A|XYE}(\omega, p_{XY})$ is increasing in ω.*

Proof. Let us fix the score ω. From the analysis in section 5.2 we know that the optimum value of δ is $\frac{R^2}{4}\cos(2\theta)$. Throughout this proof we take $\delta = \frac{R^2}{4}\cos(2\theta)$ and consider $\rho_{A'B'}$ to depend on two parameters R and θ. Let $(\mathcal{N}^*, \rho^*) \equiv \rho(R^*, \theta^*)$ be the channel and state that that solves the optimization problem for $G_{A|XYE}(\omega, p_{XY})$, i.e., such that $G_{A|XYE}(\omega, p_{XY}) = H(A|XYE)_{(\mathcal{N}^*\otimes\mathcal{I}_E)(\rho_{A'B'E}(R^*,\theta^*))}$. It suffices to show that there exists a curve $\sigma : [-1, 0] \mapsto \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_E)$, such that

1. $\sigma(0) = \rho^*$
2. $g(t) := H(A|XYE)_{(\mathcal{N}^*\otimes\mathcal{I}_E)(\sigma(t))}$ is differentiable for all $t \in [-1, 0]$.
3. $\frac{\mathrm{d}g(t)}{\mathrm{d}t}\Big|_{t=0} > 0$
4. $\forall t : \frac{\mathrm{d}}{\mathrm{d}t} S\left((\mathcal{N}^* \otimes \mathcal{I}_E)(\sigma(t))\right) > 0.$

Then, if 1–4 hold, using the fact that $g(t)$ is continuous and has a positive derivative at $t = 0$, there exists $t_0 < 0$ such that for $t \in (t_0, 0)$, $g(t) < g(0)$. Since the $S\left((\mathcal{N}^* \otimes \mathcal{I}_E)(\sigma(t))\right)$ is continuous function, we must have that for any $t \in (t_0, 0)$

$$\begin{aligned} H(A|XYE)_{(\mathcal{N}^* \otimes \mathcal{I}_E)(\rho^*_{A'B'E})} &> H(A|XYE)_{(\mathcal{N}^* \otimes \mathcal{I}_E)(\sigma(t))} && (5.52) \\ &\geq G_{A|XYE}\left(S((\mathcal{N}^* \otimes \mathcal{I}_E)(\sigma(t)), p_{XY}\right). && (5.53) \end{aligned}$$

Since $S((\mathcal{N}^* \otimes \mathcal{I}_E)(\sigma(t)) < \omega$ this establishes the claim.

It remains to show that there exists a function $\sigma(t)$ such that 1–4 hold. Recall from section 5.2 that we can write

$$H(A|XYE) = 1 + \sum_{x \in \{0,1\}} p_X(x) H_{\text{bin}}(g(\theta, \alpha_x)) - H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}), \tag{5.54}$$

where $g(\theta, \alpha) := \frac{1}{2}\left(1 + R\sqrt{1 + \sin(2\theta)\cos(4\alpha)}\right)$. We then set

$$\sigma(t) = \rho(R^* + \kappa t, \theta^*) \tag{5.55}$$

for some positive number κ such that $R^* - \kappa > 3/4$. Thus, $\sigma(0) = \rho^*$ and differentiability of $g(t)$ can be shown using the form (5.54). We compute the t derivative:

$$\left.\frac{\mathrm{d}g(t)}{\mathrm{d}t}\right|_{t=0} = \kappa \frac{\partial}{\partial R}\Big(H(A|XYE)_{(\mathcal{N}^* \otimes \mathcal{I}_E)(\rho_{A'B'E}(R,\theta))}\Big)\Big|_{R=R^*, \theta=\theta^*}. \tag{5.56}$$

Note that

$$\begin{aligned} \frac{\partial}{\partial R} H_{\text{bin}}(g(\theta, \alpha)) &= H'_{\text{bin}}(g(\theta, \alpha)) \frac{\sqrt{1 + \sin(2\theta)\cos(4\alpha)}}{2} \\ &\geq H'_{\text{bin}}\left(\frac{1}{2} + \frac{R}{2}(\cos(\theta) + \sin(\theta))\right) \frac{\cos(\theta) + \sin(\theta)}{2} \\ &= H'_{\text{bin}}(\lambda_0 + \lambda_1) \frac{\partial}{\partial R}(\lambda_0 + \lambda_1) \\ &= \frac{\partial}{\partial R} H_{\text{bin}}(\lambda_0 + \lambda_1), \end{aligned}$$

where we have used that $H'_{\text{bin}}(p)$ is decreasing in p for $p > 1/2$, so we take $\alpha = 0$ to obtain a bound. It follows that

$$\left.\frac{\mathrm{d}g(t)}{\mathrm{d}t}\right|_{t=0} = \kappa \frac{\partial}{\partial R}\Big(H_{\text{bin}}(\lambda_0 + \lambda_1) - H(\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\})\Big) > 0,$$

where the inequality is Lemma 35.

Finally, the function

$$S((\mathcal{N}^* \otimes \mathcal{I}_E)(\sigma(t)) = \frac{1}{2}\sum_{i,j} \epsilon_{ij}$$

increases linearly with t (the score is linear in R). □

Lemma 37. *For $\omega \in (\frac{3}{4}, \frac{1}{2}(1+\frac{1}{\sqrt{2}}))$, the function $G_{AB|X=0,Y=0,E}(\omega)$ is increasing in ω.*

Proof. The proof follows the same lines as the previous lemma but with the entropy changed. From section 4.6 we have

$$H(AB|X=0,Y=0,E) = 1 + H_{\text{bin}}(2\epsilon_{00}) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\}).$$

We have

$$\begin{aligned}\frac{\partial}{\partial R}H_{\text{bin}}(2\epsilon_{00}) &= H'_{\text{bin}}(2\epsilon_{00})\frac{\cos(\theta)\cos(2(\alpha_0-\beta_0))+\sin(\theta)\cos(2(\alpha_0+\beta_0))}{2} \\ &\geq H'_{\text{bin}}\left(\frac{1}{2}+\frac{R}{2}(\cos(\theta)+\sin(\theta))\right)\frac{\cos(\theta)+\sin(\theta)}{2} \qquad (5.57)\end{aligned}$$

and the remainder of the argument matches the previous proof. □

Lemma 38. *For $\omega \in (\frac{3}{4}, \frac{1}{2}(1+\frac{1}{\sqrt{2}}))$ and any distribution p_{XY}, the function $G_{AB|XYE}(\omega, p_{XY})$ is increasing in ω.*

Proof. The proof for this again follows those above, except in this case (see section 4.7)

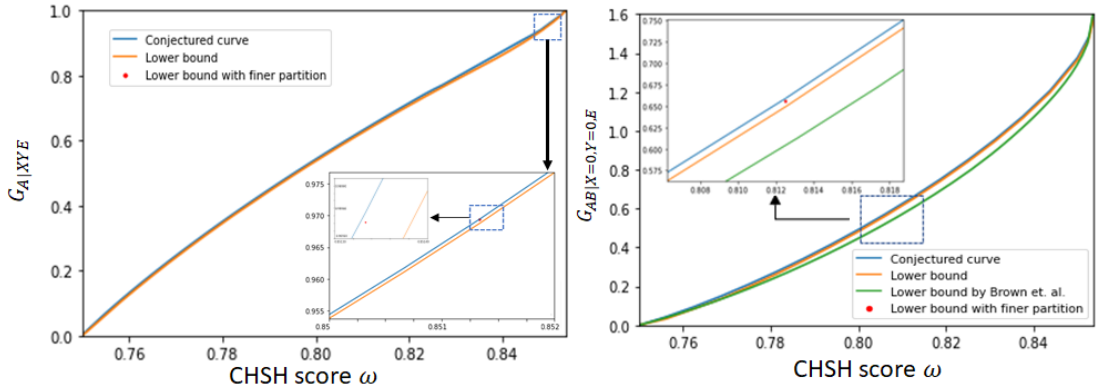
$$H(AB|XYE) = 1 + \sum_{xy} p_{XY}(x,y)H_{\text{bin}}(2\epsilon_{xy}) - H(\{\lambda_0,\lambda_1,\lambda_2,\lambda_3\}). \quad (5.58)$$

The bound that holds for ϵ_{00} in (5.57) holds for all ϵ_{xy}, and hence the rest of the argument goes through as before. □

5.6 RESULTS FOR THE LOWER BOUNDS

Lower bounds generated in this way are shown in Fig. 5.1, and can be seen to be close to the upper bounds. In Fig. 5.1(b) we also compare with a lower bound on $G_{AB|X=0,Y=0,E}$ from [43]. The lower bounds from our technique can be improved by refining the partition of the domain at the expense of increasing the computational time required. As seen in Fig. 5.1(b), refining the partition moves the lower bound closer to the upper bound. The lower bounds for $F_{AB|XYE}$ can be seen in Figure 5.2 Our numerical evidence suggests that the upper-bounds generated in the previous chapters are tight, as the lower bounds appear to converge to the upper bounds. So, in the remainder of the work, we use upper bounds for computing rates for protocols.

Conjecture 1. *The upper bounds in Lemmas 20 and 21 are tight.*



(a) (b)

Figure 5.1: Graphs of the conjectured rates and lower bounds for (a) $G_{A|XYE}$ (b) $G_{AB|X=0,Y=0,E}$ with uniformly chosen inputs. For $G_{AB|X=0,Y=0,E}$ we also show a lower bound from Brown et al. [43]. We also demonstrate that the lower bound for $G_{AB|X=0,Y=0,E}$ can be tightened by refining the partitioning of the domain for a specific point (due to the increased computation time, we did not do this throughout).

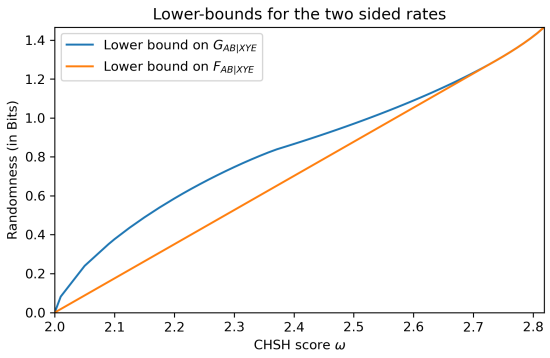


Figure 5.2: Graphs for lower bounds on $G_{AB|XYE}(\omega)$ and $F_{AB|XYE}(\omega)$

— 6 —

Results and discussion

In this chapter, we discuss CHSH-based protocols for DIRNE of both the spot-checking and non spot-checking types. We pick specific protocols for concreteness, but there are many possible variations. For instance, the protocols we discuss condense the observed statistics to a single score, but this is not necessary, and in some cases and for some sets of experimental conditions it can be advantageous to use multiple scores [51, 56].

Before getting to the protocols, we first describe the setup, assumptions and security definition. Although DIRNE requires no assumptions on how the devices used operate, the setup for DIRNE involves a user who performs the protocol within a secure laboratory, from which information cannot leak. Individual devices can also be isolated within their own sub-laboratory and the user can ensure that these devices only learn the information necessary for the protocol (in particular, they cannot learn any inputs given to other devices). The user has access to a trusted classical computer and an initial source (or sources) of trusted randomness.

The quantum devices used for the protocol are only limited by the laws of quantum theory and may share arbitrary entanglement with each other and with an adversary. However, they cannot communicate with each other, or to the adversary after the protocol starts. Furthermore, we assume they are kept isolated after the protocol (see the discussion in Appendix A.5).

For security of the protocols, we use a composable security definition. Consider a protocol with output Z and use Ω to denote the event that it does not abort. The protocol is (ϵ_S, ϵ_C)-secure if

1. $\frac{1}{2}p_\Omega||\rho_{ZE|\Omega} - \frac{1}{d_Z}\mathbb{1}_Z \otimes \rho_{E|\Omega}||_1 \leq \epsilon_S$, where E represents all the systems held by an adversary and d_Z is the dimension of system Z; and

2. There exists a quantum strategy such that $p_\Omega \geq 1 - \epsilon_C$.

Here ϵ_S is called the soundness error, and ϵ_C is the completeness error.

6.1 CHSH-BASED SPOT-CHECKING PROTOCOL FOR RANDOMNESS EXPANSION

We now describe a spot-checking protocol for randomness expansion. It uses a central biased random number generator R_T and two other random number generators, R_A and R_B that are near each of the devices used to run the protocol.

Protocol 1. (Spot-checking protocol)

Parameters:
n – number of rounds
γ – test probability
ω_{exp} – expected CHSH score
δ – confidence width for the score

1. Set $i = 1$ for the first round, or increase i by 1.
2. Use R_T to choose $T_i \in \{0, 1\}$ where $T_i = 1$ occurs with probability γ.
3. If $T_i = 1$ (test round), R_A is used to choose X_i uniformly, which is input to one device giving output A_i. Likewise R_B is used to choose Y_i uniformly, which is input to the other device giving output B_i. Set $U_i = 1$ if $A_i \oplus B_i = X_i Y_i$ and $U_i = 0$ otherwise.
4. If $T_i = 0$ (generation round), the devices are given inputs $X_i = Y_i = 0$, and return the outputs A_i and B_i. Set $U_i = \perp$.
5. Return to Step 1 unless $i = n$.
6. Calculate the number of rounds in which $U_i = 0$ occurred, and abort the protocol if this is larger than $n\gamma(1 - \omega_{\text{exp}} + \delta)$.
7. Process the concatenation of all the outputs with a quantum-proof strong extractor Ext to yield $\text{Ext}(\mathbf{AB}, \mathbf{R})$, where $\mathbf{R}$ is a random seed for the extractor. Since a strong extractor is used, the final outcome can be taken to be the concatenation of $\mathbf{R}$ and $\text{Ext}(\mathbf{AB}, \mathbf{R})$ (see Section 2.4 for details of randomness extraction).

There are a few important points to take into account when running the protocol. Firstly, it is crucial that each device only learns its own input and not the value of the other input, or of T_i. If this is not satisfied it is easy for devices to pass the protocol without generating randomness. Secondly, for implementations in which devices can fail to record outcomes when they should, it is important to close the detection loophole, which can be done by assigning an outcome, say 0, when a device fails to make a detection.
In order to run the protocol, some initial randomness is needed to choose which rounds are test rounds, to choose the inputs in the test rounds and to seed the extractor. Since the extractor randomness forms part of the final output, it is not consumed in the protocol, so for considering the rate at which the protocol consumes randomness we can work out the amount of uniform randomness needed to supply the inputs. Using the rounded interval algorithm [63] to make the biased random number generator, $n(H_{\text{bin}}(\gamma) + 2\gamma) + 3$ is the expected amount of input randomness required. To achieve expansion, the number of output bits must be greater than this. We use the entropy accumulation theorem (EAT) to lower bound the amount of output randomness. Asymptotically the relevant quantity is $H(AB|X = 0, Y = 0, E)$. The quantity $H(A|X = 0, Y = 0, E)$ acts as a lower bound for this, and can be used in its place if convenient, for instance in analyses that are more straightforward with an analytic curve.

6.2 CHSH-BASED PROTOCOLS WITHOUT SPOT-CHECKING

In this section we discuss two protocols which do not require spot checking. Protocol 2 uses two biased local random number generators to choose the inputs on each round. Protocol 3 eliminates the bias, but also recycles the input randomness. Recycling the input randomness is necessary when unbiased random number generators are used, since otherwise more randomness is required to run the protocol than is generated. Protocol 3 gives the highest randomness generation rate of all the protocols we discuss.

Protocol 2. (Protocol with biased local random number generators)

Parameters:
n – number of rounds
ζ^A – probability of 1 for random number generator R_A (taken to be below $1/2$)
ζ^B – probability of 1 for random number generator R_B (taken to be below $1/2$)

ω_{exp} – expected CHSH score.
δ – confidence widths for each score.

1. Set $i = 1$ for the first round, or increase i by 1.

2. Use R_A to choose $X_i \in \{0, 1\}$, which is input to one of the devices giving output $A_i \in \{0, 1\}$. Likewise use R_B to generate $Y_i \in \{0, 1\}$, which is input to the other device giving output $B_i \in \{0, 1\}$. Here $X_i = 1$ occurs with probability ζ^A and $Y_i = 1$ occurs with probability ζ^B. Set $U_i = (X_i, Y_i, 1)$ if $A_i \oplus B_i = X_i Y_i$ and $U_i = (X_i, Y_i, 0)$ otherwise.

3. Return to Step 1 unless $i = n$.

4. Compute the value

$$\omega = \frac{1}{4} \sum_{x,y} \frac{|\{i : U_i = (x, y, 1)\}|}{n p_X(x) p_Y(y)} \tag{6.1}$$

and abort the protocol if $\omega < \omega_{\text{exp}} - \delta$. Here $p_X(1) = \zeta^A$, $p_X(0) = 1 - \zeta^A$, $p_Y(1) = \zeta^B$ and $p_Y(0) = 1 - \zeta^B$.

5. Process the concatenation of all the outputs with a quantum-proof strong extractor Ext to yield $\text{Ext}(\mathbf{AB}, \mathbf{R})$, where $\mathbf{R}$ is a random seed for the extractor. Since a strong extractor is used, the final outcome can be taken to be the concatenation of $\mathbf{R}$ and $\text{Ext}(\mathbf{AB}, \mathbf{R})$ (see Section 2.4 for details).

Note that the quantity $|\{i : U_i = (x, y, 1)\}|/(n p_X(x) p_Y(y))$ in (6.1) is an estimate of the probability of winning the CHSH game for inputs $X = x$ and $Y = y$, and hence the ω computed in Step 4 is an estimate of the CHSH value that would be observed if the same setup was used but with X and Y chosen uniformly.
The input randomness required per round in this protocol is roughly $H_{\text{bin}}(\zeta^A) + H_{\text{bin}}(\zeta^B)$. To quantify the amount of output randomness (before randomness extraction is performed), in the asymptotic limit similar to the spot checking protocol, the relevant operational quantity is the von Neumann entropy $H(AB|XYE)$. Expansion hence cannot be achieved if $H(AB|XYE) - H_{\text{bin}}(\zeta^A) - H_{\text{bin}}(\zeta^B) < 0$, which places constraints on the pairs of possible (ζ^A, ζ^B). For ζ^A and ζ^B smaller than $1/2$, the quantity $H(AB|XYE) - H_{\text{bin}}(\zeta^A) - H_{\text{bin}}(\zeta^B)$ increases as ζ^A and ζ^B decrease, and hence we want to take these to be small. They only need to be large enough to ensure that $X = 1, Y = 1$ occurs often enough to give a good estimate of the empirical score.

Since

$$\begin{aligned} H(AB|XYE) &= \sum_{xy} p_{XY}(x,y) H(AB|XYE) \\ &\geq \min_{x,y} H(AB|X=x, Y=y, E), \end{aligned} \tag{6.2}$$

we can use the bounds formed for $H(AB|X=0,Y=0,E)$ instead, albeit with a loss of entropy (this loss of entropy is small if ζ_A and ζ_B are small)[1].

One reason for using Protocol 2 rather than Protocol 1 is that the former enables the locality loophole to be closed while expanding randomness. In order to perform the Bell tests as part of a device-independent protocol we need to make inputs to two devices in such a way that neither device knows the input of the other. One way to ensure this is by using independent random number generators on each side of the experiment, and ensuring the outcome of each device is given at space-like separation from the production of the random input to the other. Although space-like separation can provide a guarantee (within the laws of physics) that each device does not know the input of the other[2], in a cryptographic setting it is necessary to assume a secure laboratory to prevent any unwanted information leaking from inside the lab to an eavesdropper. The same mechanism by which the lab is shielded from the outside world can be used to shield devices in the lab from one another and hence can prevent communication between the two devices during the protocol. However, although unnecessary for cryptographic purposes, it is interesting to consider closing the locality loophole while expanding randomness.

This is not possible in a typical spot-checking protocol, where a central random number generator is used to decide whether a round is a test round or not. Considering Protocol 1, the locality loophole can be readily closed during the test rounds, but the use of the central random number generator means that, if one is worried that hidden communication channels are being exploited, there is a loophole that the devices could behave differently on test rounds and generation rounds. For instance, measurement devices that know whether a round is a test or generation round could supply pre-programmed outputs in generation rounds, while behaving honestly in test rounds. Thus, spot-checking protocols do not enable fully closing the locality loophole while expanding randomness.

[1] There is nothing special about the choice $X=0$ and $Y=0$ when computing the bounds for $H(AB|X=0,Y=0,E)$.

[2] Provided we have a reasonable way to give a time before which the output of R_A and R_B did not exist.

When using Protocol 2 with $\zeta^A = \zeta^B - \zeta$, the main difference to Protocol 1 is that the distribution of X and Y is $((1-\zeta)^2, \zeta(1-\zeta), \zeta(1-\zeta), \zeta^2)$ rather than $(1 - 3\gamma/4, \gamma/4, \gamma/4, \gamma/4)$. In the analysis this manifests itself in the statistics, and the much lower probability of $X = 1, Y = 1$ requires an adjustment of δ to achieve the same error parameters for the protocol. A comparison between the output rates for Protocols 1 and 2 is shown in Figures 6.1 and 6.2.

Protocol 3. (Protocol with recycled input randomness)

Parameters:
n – number of rounds
ω_{exp} – expected CHSH score.
δ – confidence width.

1. Set $i = 1$ for the first round, or increase i by 1.
2. Use R_A to choose $X_i \in \{0, 1\}$ uniformly, serving as the input to one of the devices giving output $A_i \in \{0, 1\}$. Likewise use R_B to generate $Y_i \in \{0, 1\}$ uniformly, which is input to the other device giving output $B_i \in \{0, 1\}$. Set $U_i = 1$ if $A_i \oplus B_i = X_i Y_i$ and $U_i = 0$ otherwise.
3. Return to Step 1 unless $i = n$.
4. Count the number of rounds for which $U_i = 0$ occurred and abort the protocol if this is above $n(1 - \omega_{\text{exp}} + \delta)$.
5. Process the concatenation of all the inputs and outputs with a quantum-proof strong extractor Ext to yield $\mathrm{Ext}(\mathbf{ABXY}, \mathbf{R})$, where $\mathbf{R}$ is a random seed for the extractor. Since a strong extractor is used, the final outcome can be taken to be the concatenation of $\mathbf{R}$ and $\mathrm{Ext}(\mathbf{ABXY}, \mathbf{R})$.

An important difference in this protocol compared to Protocols 1 and 2 is in the extraction step, which now extracts randomness from the input strings $\mathbf{X}$ and $\mathbf{Y}$ as well as the outputs. Without recycling the inputs, expansion would not be possible in Protocol 3. With this modification, the relevant quantity to decide the length of the output is $H(ABXY|E)$, and so $H(AB|XYE) = H(ABXY|E) - H(XY) = H(ABXY|E) - 2$ is the relevant quantity for calculating the rate of expansion. Note that in order to reuse the input in a composable way, it also needs to be run through an extractor [2] (for a

discussion of why it is important to do so and a few more composability-related issues, see Appendix A.5).

We could also consider an adaptation of Protocol 1 in which the input randomness is recycled, forming Protocol 1′ from Protocol 1 by replacing Step 7 by

7′. Process the concatenation of all the inputs and outputs with a quantum-proof strong extractor Ext to yield $\mathrm{Ext}(\mathbf{ABXY}, \mathbf{R})$, where $\mathbf{R}$ is a random seed for the extractor. Since a strong extractor is used, the final outcome can be taken to be the concatenation of $\mathbf{R}$ and $\mathrm{Ext}(\mathbf{ABXY}, \mathbf{R})$.

In this case, as the number of rounds, n, increases the advantage gained by this modification decreases, becoming negligible asymptotically. This is because as n increases, the value of γ required to give the same overall security tends to zero, and hence the amount of input randomness required becomes negligible. Note that recycling the input randomness in Protocol 2 in the case where $\zeta^A = \zeta^B = 1/2$ is equivalent to Protocol 3. Like Protocol 2, Protocol 3 also allows the locality loophole to be closed if on each round i, the random choice X_i is space-like separated from the output B_i and the random choice Y_i is space-like separated from the output A_i.

In each of the protocols, the parameter δ should be chosen depending on the desired completeness error. For the spot-checking protocol, the relation between the two is discussed in [31, Supplementary Information I D]. The analysis there can be applied to the protocol with recycled input randomness by setting $\gamma = 1$ and the protocol with biased local random number generators is discussed in Appendix A.4.3.2.

Figures 6.1 and 6.2 show how the amount of certifiable randomness varies with the score, ω, and round number, n. Note that in the cases where the rate curves are linear, they are linear for most of their ranges. Extending the linear part to the full range of quantum scores makes it easier to use the EAT while only resulting in a small drop in rate for scores close to the maximum quantum value. Note that, as mentioned above, strictly speaking, the numerical curves we provided for the von Neumann entropy are upper bounds; the curves in Figures 6.1 and 6.2 are generated under the assumption that these upper bounds are tight. [We could also use our lower bound instead. This would result in a small down-shifting of the curves, but means that the bounds are provably reliable.]

To demonstrate the increased practicality of the two-sided curves, we use the parameters from a recent experiment [31] with Protocol 1. There a score of just over 0.752 was obtained, for which it would require about 9×10^{10} rounds to achieve expansion using Protocol 1 with $\gamma = 3.383 \times 10^{-4}$, $\epsilon_S = 3.09 \times 10^{-12}$, $\epsilon_C = 10^{-6}$ and taking the one-sided

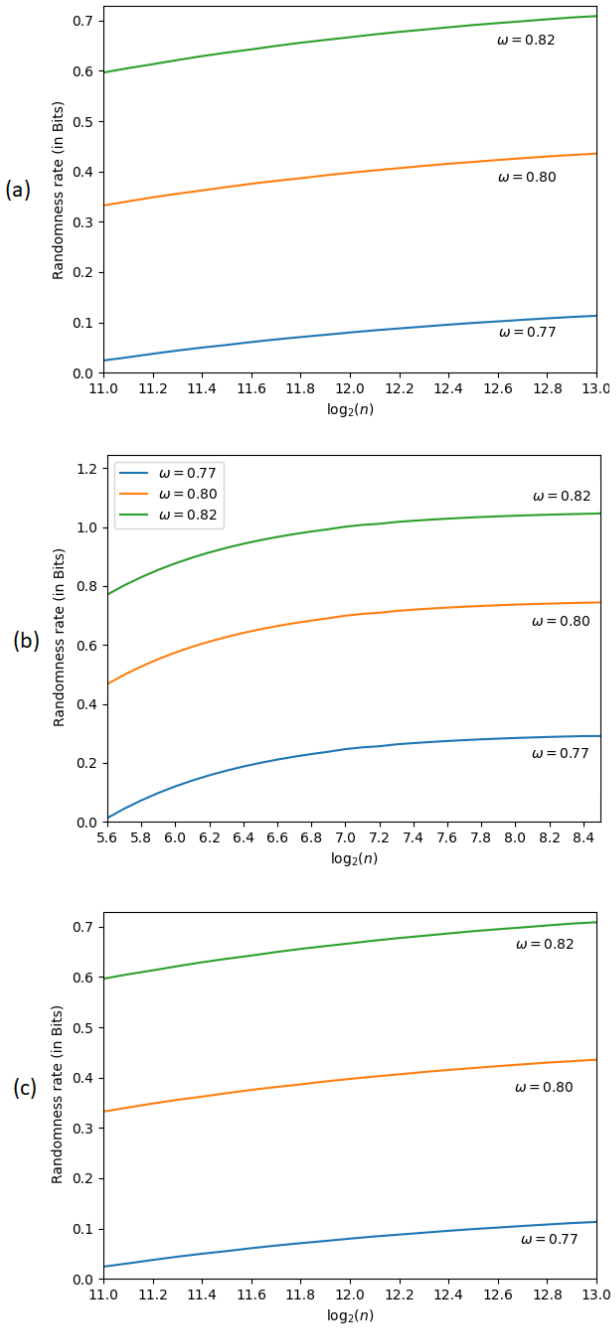


Figure 6.1: Graphs of the net rate of certifiable randomness according to the EAT for (a) the spot checking protocol (Protocol 1), (b) the protocol with recycled input randomness (Protocol 3), and (c) the protocol with biased local random number generators (Protocol 2), showing the variation with the number of rounds for three different scores, ω. The error parameters used were $\epsilon_S = 3.09 \times 10^{-12}$ and $\epsilon_C = 10^{-6}$. For each point on the curve (a) an optimization over γ was performed to maximize the randomness; similarly, the values of $\zeta^A = \zeta^B$ were optimized over to generate the curves in (c).

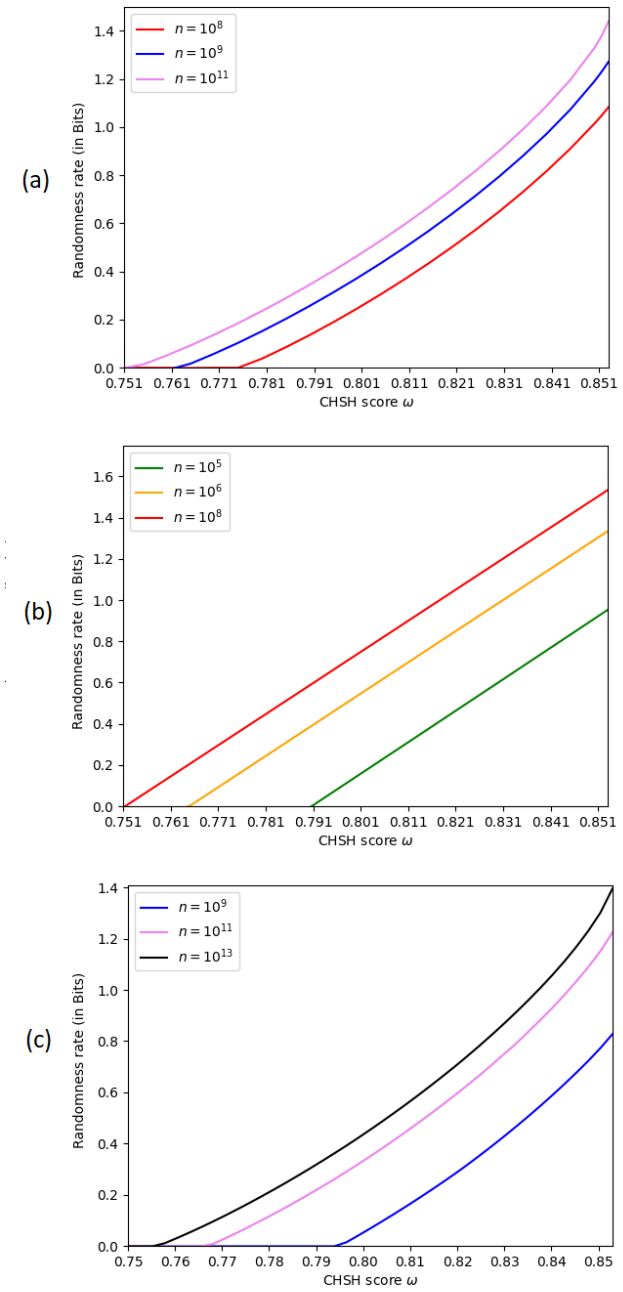


Figure 6.2: Graphs of the net rate of certifiable randomness according to the EAT for (a) the spot checking protocol (Protocol 1), (b) the protocol with recycled input randomness (Protocol 3), and (c) the protocol with biased local random number generators (Protocol 2), showing the variation with the CHSH score ω. The round numbers, n, are indicated in the legend. The error parameters used were $\epsilon_S = 3.09 \times 10^{-12}$ and $\epsilon_C = 10^{-6}$. As in Figure 6.1, the values of γ (for (a)) and $\zeta^A = \zeta^B$ (for (c)) were optimized over for each point.

randomness [31]. Using Protocol 3 instead, and taking the two-sided randomness for the same score and error parameters allows expansion for $n \gtrsim 8 \times 10^7$, significantly increasing the practicality. For instance, the main experiment of [31] was based on a spot-checking protocol and took 19.2 hours; the use of Protocol 3 instead would allow the same amount of expansion in about 60 seconds (this time holds under the assumption that the same repetition rate of the experiment can be met in the non-spot checking protocol[3]). Protocol 2, however, produces lower randomness rates compared to the spot-checking protocol. This is partly because more input randomness is required, and also because the completeness error has a worse behaviour. Protocol 2 is hence useful when inputs are not recycled and when closing the locality loophole is desirable.

When discussing randomness expansion we have considered the figure of merit to be the amount of expansion per entangled pair shared. An alternative figure of merit is the ratio of the final randomness to the initial randomness, i.e., here we are considering how much randomness we can get from a given amount of initial randomness. For the latter figure of merit, Protocol 3 is no longer optimal, since the amount of expansion cannot exceed the amount of input randomness. For the other two protocols the ratio of output randomness to input randomness can be made much higher by taking either γ or $\zeta^A\zeta^B$ to be small.

6.3 DISCUSSION

In chapters 3, 4, 5, and 6 we discussed CHSH based protocols for randomness expansion. We have given numerical bounds on various conditional von Neumann entropies that are relevant for CHSH-based device-independent protocols and discussed when each can be applied. We have investigated their implications using explicit protocols, comparing the finite statistics rates using the EAT, showing use of two-sided randomness has the potential to make a big difference. We also looked at protocols beyond the usual spot checking type. The first removes the spot checking to allow expansion while closing the locality loophole, and the second recycles the input randomness, so allowing expansion while performing a CHSH test on every round.

It remains an open question to find an analytic form for $F_{AB|X=0,Y=0,E}$, $F_{A|E}$ and $F_{AB|E}$. Since the curves $F_{A|E}$ and $F_{AB|E}$ are linear for all but the very highest (experimentally

[3] In some experiments, the rate at which we can switch between the two measurements is relatively slow, and hence when using Protocol 3, where switching is required on most rounds, the switching rate dominates, slightly increasing the time.

least achievable) scores, in these cases not much is lost by extending the line to all scores forming a lower bound that tightly covers all of the experimentally relevant cases. On the other hand $F_{AB|X=0,Y=0,E}$ is a convex curve throughout and hence a tight analytic form would be particularly useful in this case. Our initial analysis suggests that the form of the parameters achieving the optimal values for these functions is sufficiently complicated that any analytic expression would not be compact. A reasonably tight analytic lower bound for $F_{AB|X=0,Y=0,E}$ could also be useful for theoretical analysis. Note also that the bound $F_{A|E} \geq F_{A|XYE}$ appears to be fairly tight (see Figure 4.1(a)) so the analytic form for $F_{A|XYE}$ can be used to bound $F_{A|E}$ with little loss. Another open problem is to find a concrete scenario in which $F_{AB|E}$ is directly useful.

The use of Jordan's lemma in this work prevents the techniques used being extended to general protocols, and finding improved ways to bound the conditional von Neumann entropy numerically in general cases remains of interest. For example, protocols that use three inputs for one party can allow up to 2 bits of randomness per entangled pair (see, e.g. [51]), and a way to tightly lower bound the von Neumann entropy in this case would further ease the experimental burden required to demonstrate DIQKD in the lab.

Part II

Semi Device Independent Protocols

— 7 —

Introduction to semi-Device Independent Protocols

7.1 INTRODUCTION

As outlined in the introduction, the Device Independent protocols of randomness expansion make minimal assumptions for certifying randomness. However, as they currently stand, these protocols are extremely difficult to implement in a laboratory setting, even with the most advanced technology available. The primary obstacle in this regard is the performance of a loophole-free Bell test. There are two main loopholes that have traditionally caused problems in performing the Bell test. The first is the locality loophole, which refers to the requirement that the devices are spacelike separated when carrying out the Bell test. The second is the detection loophole, resulting from inefficient measurements.

For protocols of randomness expansion, the locality loophole might not be as critical to close, assuming we have control over the laboratory. Shielding mechanisms can be used between the devices to ensure that no communication occurs among them. Theoretically, this could be the same shielding that prevents any data leak from the lab to the outside world. The remaining challenge is the detection loophole, the closure of which requires very high-efficiency detectors. While such detectors are available, they tend to be expensive, and maintaining good-quality entanglement over a relatively large distance presents a significant practical problem.

As noted in the introduction, progress is being made in this direction, but until these challenges can be overcome in a more practical way, Device Independent protocols for randomness expansion can only be used for a select few top-security applications. On a

practical level, however, the security of random numbers is desired. Therefore, there is a need for protocols that are not fully Device Independent but rather allow for a certain degree of trust on the components.

By imposing well-chosen assumptions on the devices or the system's underlying process, semi-Device Independent (semi-DI) approaches can achieve most of the security benefits of a fully Device Independent (DI) protocol whilst bypassing the need for challenging experimental implementations. This intermediate scenario strikes a balance between security and practicality, making it a promising approach for implementing quantum random number generators.

Many semi-DI protocols have been introduced in the scientific literature. A typical setup for these protocols involves having a source that prepares a quantum state, and then sends it to a measurement device that performs a measurement on the prepared quantum state. One advantage of this method is that it doesn't rely on sharing entangled states over a distance, which is a requirement of the Device Independent protocols. This type of experimental setup is often referred to as a "prepare and measure" setup (or scenario).

The assumptions underlying a semi-DI scheme can vary based on specific needs. For example, one might consider source-DI or measurement-DI cases, where either the source device or the measurement device is trusted, respectively.

In this section of the thesis, we work on protocol based on a "prepare and measure" scenario where the source and measurement devices are both uncharacterized. Thus, it is both source-DI and measurement-DI. This method was first introduced in [9] and been subsequently been studied in [64–69]. These works discuss semi-DI Quantum Random Number Generators (QRNGs) that employ a physical system with a unique ground state – i.e. the lowest eigenvalue state of the system Hamiltonian is unique. The outlined protocols are based on the source preparing the states with low energy or high overlap[1] with the unique ground state. The component of the protocol responsible for the verification of these energy and overlap constraints is trusted, hence making these protocols semi-Device Independent.

The idea for the protocol is as follows: Recall from section 2.2 that two quantum states ρ^0 and ρ^1, when prepared with a uniform probability distribution, cannot be perfectly distinguished if the distance between them, represented as $||\rho^0 - \rho^1||_1$, is small. Now consider the following scenario: we randomly select $X \in \{0, 1\}$ and prepare ρ^x ($x = 0$ or 1) depending on the outcome of $X = x$, and then task our measurement device with

[1] Overlap here is defined as the fidelity between the state and the ground state.

distinguishing between the states ρ^0 and ρ^1 to produce a bit Y with the intent that $Y = X$. If the states ρ^0 and ρ^1 are sufficiently close[2], then the probability that $Y = X$ is strictly less than 1, insinuating that Y must possess randomness, even if X is later revealed. This forms the basis of our protocol.

The advantage of such a protocol is that it has the features of Device Independent protocols, in the sense that the security of the protocol solely relies on the input-output statistics, namely the probability that the outputs and inputs are different - i.e. $P(Y \neq X)$. Furthermore, observe that the protocol does not also depend upon which states are prepared, all that matters is the quantity $||\rho^0 - \rho^1||_1$. Intuitively, given that the protocol hinges on the inability to perfectly distinguish sufficiently close quantum states, one would anticipate that the randomness generated in the protocol increases as the distance between the states decreases. This intuition will be confirmed when we compute the rates of the protocols in Chapter 9. Thus, if there is an experimental way to ensure the distance $||\rho^0 - \rho^1||_1$ is small, then we have a good protocol for randomness expansion.

To ensure the distance between the generated states is small, one strategy involves seeking systems with a unique ground state, meaning the system's lowest energy state is non-degenerate. If the system's energy is observed to approximate the vacuum energy (i.e., the energy of the ground state), then both states must be nearly equivalent to the ground state and, by extension, to each other. Provided that the measured energy is low enough, it is possible to find bounds on the distance between the states simply by knowing the energies of the states produced using simple arguments. Note that theoretically, the ground state isn't inherently special; the key lies in generating nearly identical states. In practical terms, when the state source is a laser pulse - common in many prepare-and-measure situations - the source might produce two coherent states, $|\alpha_0\rangle\langle\alpha_0|$ and $|\alpha_1\rangle\langle\alpha_1|$, both with minimal mean photon numbers, rendering them close to the vacuum state.

Unfortunately, measuring the energy of the system is not possible in a Device Independent fashion. Thus, we need to have a trusted power meter, whose role is to measure the energy of the states, making this protocol semi-Device Independent. We use the same principle for our protocol of randomness, however, we do not measure energy of the states. Rather, we assume that we have access to a power-meter, which can determine if the prepared states have more than a threshold energy - so it is a yes/no machine (similar to an on/off photo diode). The details of this will become clear in the next section. We

[2]Here, two states are close to each other if the trace distance between them is small.

can also include an additional component such as a variable attenuator, that can help by reducing the energy of the emitted states.

For this semi-DI protocol, the primary assumption is that the power meter is a trusted component and has not been tampered with by any adversary nor is it damaged. However, depending on the context, the studies mentioned above have required one or more additional assumptions:

- Each protocol round is independent of the previous one, and the initial conditions are identical before each round (i.i.d. assumption). This condition implies that the source or the measurement devices do not have an internal memory set by the eavesdropper.
- Only classical side-information is considered, which can be caused by device imperfections or classical correlations. This assumption limits the eavesdropper from being entangled with the quantum state prepared by the source. Furthermore, this also supposes that the eavesdropper has no quantum memory.
- The source and measurement devices are not entangled with each other.

The assumptions above can be rather strong in many scenarios. For instance, the assumption that all the rounds are identical and that the eavesdropper only has a classical side information is difficult to verify in the experimental setting. The above assumptions also explicitly rule out the case, in which, the adversary shares entanglement with the source and the measurement device, which cannot be easily justified when the RNG is purchased from a untrustworthy party. Furthermore, device imperfections, environmental changes, and experimental errors make it impossible to achieve identical experiment rounds. Furthermore, suppose the adversary has a quantum correlation with the devices, such as generated states being entangled with her states. In that case, it becomes easier for her to predict the QRNG measurement outcome.

Our analysis discards all these assumptions, except for the last one, in which the source and the measurement devices share some entanglement. Though we do allow for pre-shared randomness between the source and the measurement device. We permit the eavesdropper to share entanglement with the source and the states prepared by the source. We also allow for a fully uncharacterized measurement device known to the adversary and permit the eavesdropper, source, and measurement device to pre-share arbitrary classical randomness, thereby accounting for a quantum adversary.

In quantum state preparation and measurement, the memory effect means that earlier measurements can influence subsequent ones, and earlier prepared states may impact

those prepared later. These memory effects have implications for security, as they introduce correlations and dependencies in the outcomes. To mitigate this issue, we consider the memory effect and other potential sources of correlation in the security estimation stage.

The main advantage is that our protocol is formulated in a way that the Entropy Accumulation Theorem (EAT) can be readily applied. Though this has not been done in the thesis, it can be done in a relatively straightforward manner, thus exploiting several benefits of EAT. The primary advantage is the relaxation of the i.i.d. assumption in the protocol, and accounting for memory effects. Furthermore, the EAT also accounts for the adversary holding quantum side information. Using EAT, the problem of computing randomness rates (randomness per round) is reduced to computing the lower bound on the single round von Neumann entropy of a representative round of a protocol. The problem of computing lower bounds on the single round von Neumann entropies for the semi-DI protocol described above is one of the main aims of this section of the thesis.

One of the challenging aspects of the protocol is that no assumptions have been made on the dimensions of the states ρ^x. Fortunately, similar to DI protocols, we show that we can leverage Jordan's Lemma for such semi-DI protocols, significantly reducing the complexity of computing the rates of the protocols. Jordan's Lemma allows us to relax the problem to a scenario where the source generates qubit states, and the measurement device performs projective measurements, simplifying the process into an optimization problem involving less than eleven variables. Further simplifications, along with reliable numerical techniques for obtaining lower bounds on polynomial optimization problems as discussed in Chapter 2, enable us to reliably compute the rates for the protocol. The problem of computing these rates will be discussed in the next chapter (Chapter 8).

In this work, we present two types of protocols: one for randomness expansion, similar to protocol 3 in the DI setting, where input randomness is recycled, and another for converting public randomness to private randomness, analogous to protocol 2 in the DI setting. The protocols are detailed in Chapter 9, along with a discussion on the asymptotic rates for these protocols.

In the remainder of the chapter, we do a brief literature survey of different Quantum Random Number Generators (QRNGs) and then proceed to giving a sketch of the protocol that we study in this thesis.

7.2 A BRIEF REVIEW OF DIFFERENT QRNGS

Before delving into the main semi-DI protocol of randomness expansion discussed in this thesis, we deliver a very brief and non-exhaustive survey of the literature on various Quantum Random Number Generators (QRNGs). There is a vast literature on a range of different protocols for building QRNGs. As discussed in the previous section, there is generally a trade-off between the ease of experimental implementation of a protocol and the security of the protocol. Here, we call a protocol more "secure" if it requires fewer experimental assumptions. Broadly speaking, QRNGs fall into one of these categories:

1. Fully Device-dependent protocols (DD) (for example see [70–72]);
2. Source DI protocols (for example see [73, 74]), wherein the source is untrusted but the measurement apparatus is reliable;
3. MDI protocols (for example see [75–77]), where the measurement device is untrusted, but not the source;
4. Semi-DI protocols;
5. DI protocols.

The DD protocols are the easiest to implement since all components are trusted, and no characterization of the device is needed. However, this simplicity might come at the expense of security. On the other hand, the DI protocols are the most difficult to implement with the current technology but promise the highest possible security. The source-DI, measurement-DI, and semi-DI protocols are in the middle with security somewhere in between DD and DI and the practicality also somewhere in between DD and DI. Source and measurement DI protocols have also been studied widely in the literature. These protocols are useful when the user can either characterize the source, which can prepare desired states with minimal noise, or characterize the measurement device that comes equipped with highly efficient detectors.

The exploration of semi-DI protocols in the quantum information theory literature can be traced back to Liang et al. [78]. Their work focused on semi-DI protocols for entanglement detection. Building upon these concepts, Pawłowski et al. [79] extended this framework to include semi-DI protocols for Quantum Key Distribution (QKD). This work was further extended by Li et al. [80] to construct a semi-DI protocol for randomness expansion. These semi-DI protocols operate within the prepare-and-measure scenario and do not make any assumptions about the internal workings of the device, other than

the assumption that the dimension of the Hilbert space of the produced states is both bounded and known.

The primary challenge with randomness expansion protocols that depend on a dimension bound for the prepared states is the inability to experimentally verify such an assumption. While a lower bound on the Hilbert space dimension can be certified without trusting the devices [81, 82], an upper bound cannot be similarly certified. This limitation stems from the fact that the dimension of the Hilbert space is not a directly measurable quantity in quantum theory, leaving no reliable method to validate assumptions about the Hilbert space dimensions of prepared states.

The protocol by Van Himbeeck et al. addresses this issue [9]. Their idea is that the dimension of the Hilbert space can be indirectly deduced through energy measurements. In many quantum systems, both the Hamiltonian and its spectrum are known. The essence of the protocol is that the Hilbert space dimension can be inferred by measuring the energy of the system. For example, if the source prepares quantum states in a system with a known, non-degenerate energy spectrum $\{E_i\}_{i=0}^{\infty}$, and the highest observed energy after numerous measurements is E_k, we can deduce, with high confidence, that the Hilbert space dimension is $k+1$. This method remains applicable to a degenerate spectrum, as long as the full spectrum, including degeneracies, is known. Importantly, Van Himbeeck et al. [9] realized that if the measured energy stays below a specific threshold, the system can effectively be seen as two-dimensional, with one ground state and all other states can be effectively combined as a single excited state. Subsequent research has further developed and expanded upon these findings [64–69].

The method described above for building semi-DI protocols is by no means exhaustive. There are other semi-DI protocols based on different physical or information-theoretic principles (for example, see references [83–85]). Moreover, just as quantum behavior can be certified using Bell non-locality, another fundamental concept, contextuality, can also certify whether the experimentally observable statistics admit a quantum behaviour or not. Unfortunately, unlike non-locality, non-contextuality cannot be tested in a Device Independent manner. However, without delving into any details, if we place appropriate (partial) trust in the source or the measurement device, it is possible to devise semi-DI protocols of randomness expansion based on the distinction between contextual and non-contextual behaviors. The groundwork for such protocols has been established for the QKD setting in [86] and later extended to randomness expansion in [87].

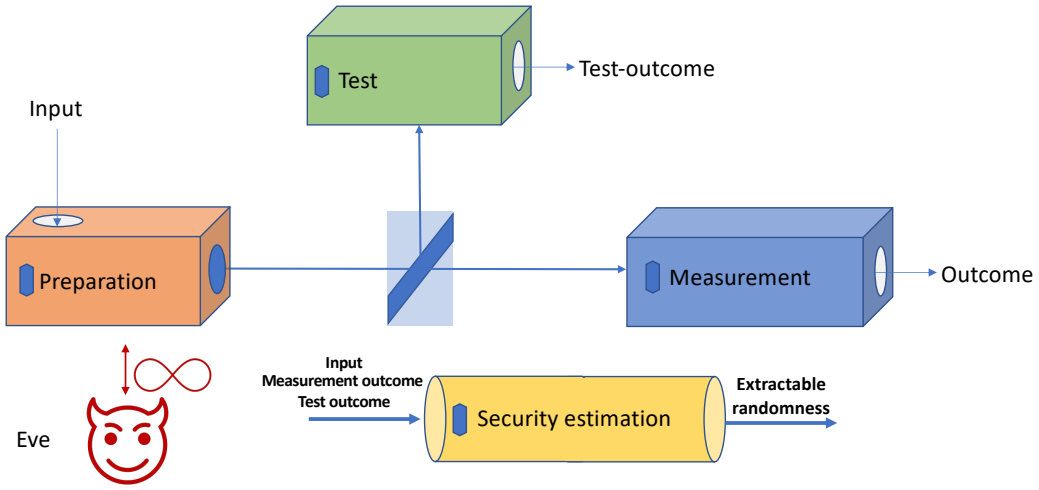


Figure 7.1: A schematic diagram of our semi-DI protocol.

7.3 GENERAL SEMI-DEVICE INDEPENDENT PROTOCOL

Figure 7.3 offers a schematic illustration of the protocol, which is split into three primary stages: preparation, testing, and measurement. It's crucial to note that full characterization is necessary only for the component in the testing phase. Both the source and measurement stages remain uncharacterized.

For the protocol to operate effectively, two separate input seeds are essential. The initial seed guides the source in state preparation, while the secondary seed dictates whether the prepared states undergo testing. Keeping these seeds hidden from potential adversaries before the protocol's initiation is vital for protocol's security .

Here is a brief overview of the protocol:

Preparation Phase: The source accepts input of a random variable $X \in \{0, 1\}$ that is generated with a probability distribution $p_X(x)$. Based on this input, the source prepares either ρ^0 or ρ^1 contingent on whether X equals 0 or 1. The state then gets relayed to component BS, which determines if the state should proceed to the testing or measurement phase. Essentially, device BS functions as a switch. Utilizing a random number T produced by the random number generator R_γ, BS sends the signal to either testing or measurement (for example, if it receives $T = 0$, then it will send the signal to measurement, otherwise it sends the signal for testing). The bias of R_γ is set to ensure most signals reach the measurement phase. If BS channels the state for testing, we denote that round a test round; otherwise, we call the round ameasurement round.

In practice, if the source is a laser, the device BS can be a mechanical switch that controls a mirror. Depending on the random variable T, it directs the signal to the testing phase or the measurement phase. However, such a switch can be lossy and therefore may not very suitable for practical implementation. Alternatively, the device BS can be a half-silvered mirror or a beam splitter, which is a passive optical device that transmits the signal to the measurement device with probability $p_T(0) = \gamma$ or reflects the signal with probability $1 - \gamma$. If such a half-silvered mirror is used, it also needs to be fully characterized and assumed not be tampered with by any adversary.

Testing Phase: This is the protocol's fully characterized segment. Recall that the protocol derives its randomness if the states ρ^0 and ρ^1 are near identical. This is ensured by ensuring both states are close[3] to the system's unique ground state, symbolized by $|0\rangle\langle 0|$ or Π_0, often referred to as the vacuum state. Essentially, the idea is that if both states are reasonably close to the vacuum state (in the state space), they are close to each other (in the state space). Formally, the definition of closeness that we use here is overlap of two states with vacuum state. The overlap for state ρ with the vacuum is defines as:

$$\mathrm{tr}\left(\rho\Pi_0\right) = \langle 0|\,\rho\,|0\rangle\,. \tag{7.1}$$

From a theoretical viewpoint, measuring this overlap requires an on-off device that can execute a two-outcome measurement $\{\Pi_0, \mathbb{1} - \Pi_0\}$, with Π_0 representing the projection onto the vacuum state. Alternatively, the overlap can be deduced from the state's energy, given that this measured energy remains below a certain threshold. Therefore, experimentally, this mandates the use of a power-meter or photodiode. For simplicity, this on-off device is often termed a 'power meter', symbolized by PM. If we define the vacuum's energy as zero, this device's primary function becomes the detection of any existing positive energy.

Considering PM could receive states ρ^0 or ρ^1 depending on the source's preparation, there are two separate overlaps with the ground state in a protocol. In our context, the term "overlap" (of the protocol) is defined as the average of these individual overlaps of ρ^0 and ρ^1 with the ground state. Specifically, the overlap Θ for a protocol is:

$$\Theta = \frac{1}{2}\mathrm{tr}(\rho^0\Pi_0) + \frac{1}{2}\mathrm{tr}\left(\rho^1\Pi_0\right). \tag{7.2}$$

[3]Here closeness of two states is measured in terms of them having either high fidelity or low trace distance.

Ideally, this value should be at its maximum to ensure that the source S consistently produces states with substantial individual overlaps.

Measurement Phase: Here, states ρ^0 or ρ^1 are sent to the measurement apparatus M. This device's primary role is to determine whether ρ^0 or ρ^1 has been sent. Essentially, it should produce a bit Y aiming for Y to equal X. If the overlap is notable, then M cannot perfectly distinguish between the states, ensuring that the output Y contains private randomness even if X is later revealed.

During the generation round, a "win" is declared if Y equals X. Theoretically, the measurement device performs a two outcome POVM $\{M_0, M_1 \equiv \mathbb{1} - M_0\}$. The protocol's score is the mean probability of securing a win, defined as:

$$\omega := \frac{1}{2}\operatorname{tr}\Big(M_0\rho^0\Big) + \frac{1}{2}\operatorname{tr}\Big(M_1\rho^0\Big), \tag{7.3}$$

Here, $\operatorname{tr}(M_x\rho^x)$ is the probability of winning a generate round when $X = x$ is prepared. A high score is desirable for our protocol. As we have discussed earlier achieving a high overlap is crucial for the protocol's optimal function. Yet, even with a specific overlap, a high score remains preferable. A low score can arise from strategies like the following: the source prepares two states (ρ^x depending upon the inputs $X = x$) that are best distinguished given a fixed overlap. The measurement device can act as follows: with a very high probability, it may ignore the incoming states and output using pre-shared randomness, which is accessible to the adversary. With a smaller probability, it might perform the optimal measurement that distinguishes ρ^0 and ρ^1. Such a strategy will produce negligible private randomness, as for most rounds, some pre-shared randomness is outputted.

Various protocols can be conceived considering the process of input and output strings. Future chapters will probe two such protocols, focusing on discussions about computing the randomness rate for these protocols.

— 8 —

Optimizing the von Neumann entropy

As with the Device Independent (DI) protocols, we can use the Entropy Accumulation Theorem (EAT) to compute the randomness rate (randomness per round) of the semi-DI protocol. Recall that informally, EAT states that to compute the randomness generated in a protocol, it is sufficient to calculate the von Neumann entropy of a single round that is representative of the full protocol. As we shall see in the next chapter, in the context of the protocols considered here, it suffices to compute a lower bound on single round of von Neumann entropy conditioned on the same score ω and the same overlap Θ. The main aim of the chapter is to perform this optimization problem. To do so we make this formalism more rigorous.

8.1 STRATEGIES

We begin by defining a strategy for our semi-Device Independent protocol:

Definition 16 (Strategy)**.** Let $\rho^x_{AE} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_E)$ and $\{M_0, M_1 \equiv \mathbb{1} - M_1\}$ be a POVM. A tuple $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, M_0)$ is called a strategy.

In the definition of a strategy, we allow the states to be sub-normalized as well. However, if a strategy defines our protocol, then the states ρ^x_{AE} should be normalized. Similar to the Device Independent scenario, this strategy is chosen by the adversary Eve.
Each strategy has an associated CQ state. Let p_X be the input probability distribution, then the CQ state associated with the strategy above is given by

$$\rho_{\mathcal{C}} := \sum_x p_X(x) \left|x\right\rangle\!\left\langle x\right|_X \otimes \left|y\right\rangle\!\left\langle y\right|_Y \otimes \mathrm{tr}\left(M_y \otimes \mathbb{1}_E \rho^x_{AE}\right).$$

We introduce a short hand notation of $H(Y|XE)_{\mathcal{C}} \equiv H(Y|XE)_{\rho_{\mathcal{C}}}$ when referring to the conditional von Neumann entropy of the CQ state $\rho_{\mathcal{C}}$.
Given a strategy, it is also possible to determine the score it achieves and the overlap that shall be observed, provided the strategy is used in an i.i.d. fashion. Thus it is useful to define the following:

Definition 17 (Score of a strategy)**.** The score of a strategy $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, M_0)$ is defined as:

$$S(\mathcal{C}) = \frac{1}{2} \sum_x \mathrm{tr} \left(\rho^x_{AE} M_x \otimes \mathbb{1}_E \right). \tag{8.1}$$

Note that $M_1 = \mathbb{1} - M_0$ is implicitly assumed when defining a strategy.

Similarly, we can define the overlap of the strategy (where Π_0 is the projector onto the ground state)

Definition 18 (Overlap of a strategy)**.** The overlap of a strategy $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, M_0)$ is given by

$$\mathcal{O}_{\Pi_0}(\mathcal{C}) = \sum_x \frac{1}{2} \mathrm{tr} \left(\rho^x_{AE} \Pi_0 \otimes \mathbb{1}_E \right). \tag{8.2}$$

There are certain special classes of strategies that may be interesting to consider:

Definition 19 (Pure state Strategy)**.** A strategy $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, M_0)$ is a pure-state strategy if ρ^0_{AE} and ρ^1_{AE} are pure states.

Definition 20 (Projective Strategy)**.** A strategy $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, P_0)$ is a projective strategy if P_0 is a projection operator.

In the next chapter, it will be shown that, in the asymptotic limit, the rate of the protocol can be found by computing (or by finding an appropriate lower bound for) of the function $F_{p_X}(\omega, \Theta)$. This function is given by the optimization problem:

$$\inf_{\mathcal{C} \in \Gamma[\omega, \Theta]} H(Y|XE)_{\rho_{\mathcal{C}}},$$

where $\Gamma[\omega, \Theta]$ are the strategies that achieve a fixed score ω and have an overlap Θ. - i.e.

$$\Gamma := \{\mathcal{C} : \quad S(\mathcal{C}) = \omega, \mathcal{O}_{\Pi_0}(\mathcal{C}) = \Theta\}.$$

The optimization problem above can be simplified right-away by showing that it is sufficient to restrict to cases when the states ρ^x_{AE} are pure.

Lemma 39. *For every strategy $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, M_0)$ there exists a pure state strategy $\mathcal{C}' = (\tilde{\rho}^0_{AE}, \tilde{\rho}^1_{AE}, M_0)$ such that*

$$H(Y|XE)_{\mathcal{C}} \geq H(Y|XE)_{\mathcal{C}'}$$

Proof. Suppose $\tilde{\rho}^x_{AE}$ are not pure states, then let $\rho^x_{AEE'}$ be any purification of ρ^x_{AE}. Let $\mathcal{C}' = (\tilde{\rho}^0_{AEE'}, \tilde{\rho}^1_{AEE'}, M_0)$. Then simple computation shows that $\mathrm{tr}_{E'}\, \rho_{\mathcal{C}'} = \rho_{\mathcal{C}}$. Thus

$$H(Y|XEE')_{\mathcal{C}'} \leq H(Y|XE)_{\mathcal{C}} \tag{8.3}$$

follows from strong subadditivity of the von Neumann entropy. □

As we restrict to the set of pure states for the rest of the analysis, it shall be understood that the tuple $(\rho^0_A, \rho^1_A, M_0)$ is the short-hand for any strategy $(\rho^0_{AE}, \rho^1_{AE}, M_0)$, where ρ^x_{AE} is any purification of ρ^x_A.

8.2 INCORPORATING PRE-SHARED RANDOMNESS AND REDUCTION TO PROJECTIVE STRATEGIES

When executing the protocol, it is necessary to consider potential attacks by the eavesdropper, who may possess pre-shared randomness with the source and measurement devices. During a round of the protocol, the eavesdropper could have complete knowledge of this pre-shared randomness. Additionally, the eavesdropper may instruct the devices to prepare a state based on this pre-shared randomness and can also instruct the measurement device to perform a specific measurement depending on the value of the pre-shared randomness.

To account for such an attack, we assume that the state ρ^x_{AE} can take the following most general form:

$$\rho^x_{AE} = \sum_{\lambda} p(\lambda) \rho^{\lambda}_{A\tilde{E}} \otimes |\lambda\rangle\langle\lambda|_{\Lambda}\,, \tag{8.4}$$

where $E = \Lambda\tilde{E}$, is the system accessible to E. Here, Λ represents an additional classical register held by Eve, the source, and the measurement device. It is important to note that the system Λ does not possess any information about the input X, as it solely

represents pre-shared randomness. Since the system Λ is not transferred from the source to the device (and the power meter) during each round, the overlap constraint for the protocol is expressed using the projector $\Pi_0 \otimes \mathbb{1}_\Lambda \otimes \mathbb{1}_{\tilde{E}}$.
In the strategy involving pre-shared randomness, the measurement device acts based on the pre-shared randomness. The most general measurement operator for our protocol can be defined as follows

$$\hat{M}_0 = \sum_\lambda M_0^\lambda \otimes |\lambda\rangle\langle\lambda|_\Lambda \otimes \mathbb{1}_{\tilde{E}}, \quad \hat{M}_1 = \sum_\lambda \left(\mathbb{1} - M_0^\lambda\right) \otimes |\lambda\rangle\langle\lambda|_\Lambda \otimes \mathbb{1}_{\tilde{E}}. \tag{8.5}$$

The definition of the measurement operator has been slightly altered from what was presented in the previous section. In this context, the measurement operator can act on the system Λ, which the Eavesdropper can access. This change is simply an artefact of our short hand notation. A more accurate representation of the protocol would segregate it into three separate yet perfectly correlated random variables: Λ_S for the source, Λ_M for the measurement device, and Λ_E for the eavesdropper. The states that the source produces depend on the value of Λ_S (which is a local random variable), and the measurement operator determines the value of the local random variable Λ_M and performs a measurement $\{M_\lambda, \mathbb{1} - M_\lambda\}$, depending upon the value of $\Lambda_M = \Lambda_S = \lambda$. To streamline the notation, we merged these into a single classical register, Λ. As a consequence of this simplification, we effectively allow the measurement apparatus to perform a measurement on a system that is accessible to the eavesdropper.
To summarize the discussion above: our strategy is given by a tuple $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, \hat{M}_0)$ where ρ^x_{AE} is of the form (8.4). Meanwhile, the measurement operator $\hat{M}_0$ is of the form (8.5). The CQ state relevant to the most generalized strategy is given by

$$\rho_{\mathcal{C}} = \sum_{\lambda=1}^{n} \sum_{x,y} p(\lambda) p_X(x) \, |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \mathrm{tr}_A \left(M_y^\lambda \otimes \mathbb{1}_E \rho^{x,\lambda}_{A\tilde{E}}\right) \otimes |\lambda\rangle\langle\lambda|_\Lambda \, . \tag{8.6}$$

The score and overlap of a strategy are defined via:

$$\begin{aligned} \omega(\mathcal{C}) &= \frac{1}{2} \sum_x \mathrm{tr}\left(\hat{M}_x \rho^x_{AE}\right) \\ \Theta_{\Pi_0}(\mathcal{C}) &= \frac{1}{2} \sum_x \mathrm{tr}\left(\Pi_0 \otimes \mathbb{1}_E \rho^x_{AE}\right) . \end{aligned}$$

The sets $\Gamma[\omega, \Theta]$ are defined identically as in the previous section.
We now argue that we can assume the measurement $\{M_y^\lambda, \mathbb{1} - M_y^\lambda\}$ to be a projective measurement without losing any generality. This is because any effect, M_0^λ, can be

expressed in terms of convex combination of extremal effects. In quantum theory, these extremal effects are projections [1]. Therefore, we can write

$$M_0^\lambda = \sum_\mu q_\mu M_0^{\lambda,\mu}, \quad M_1^\lambda = \sum_\mu q_\mu M_1^{\lambda,\mu},$$

for some probability distribution $\{q_\mu\}$. This means, if there are strategies that use non-projective measurements, they can be executed using an extra classical register that only the measurement device can access. Note that having this extra register does not change the score and the overlap of the strategy.

Because of the strong subadditivity of the conditional von Neumann entropy, letting Eve and the source use this extra classical register cannot worsen the strategy from the point of view of Eve. Considering that the measurement device is not fully characterized, Eve might also have this register in the form of pre-shared randomness instead. Therefore, this classical register can also be integrated into the register Λ.

We now show that the problem for computing the rate in the most general attack can be reduced to the problem of computing the rate when the eavesdropper does not pre-share any randomness. To facilitate this, we denote the set $\mathfrak{C}_1^P[\omega,\Theta] \subset \Gamma[\omega,\Theta]$ to be the set of all projective strategies which do not allow for any pre-shared randomness , i.e. the strategies for which $E = \tilde{E}$. In the following, convenv(.) represents the convex envelope (or convex lower bound) as introduced in Chapter 4 (see 8.10 for a more detailed explanation).

Lemma 40. *$F_{p_X}(\omega,\Theta) = \text{convenv}\Big(G_{p_X}(\omega,\Theta)\Big)$, where the function $G_{p_X}(\omega,\Theta)$ is computed using the optimization problem*

$$\inf_{\mathcal{C}\in\mathfrak{C}_1[\omega,\Theta]} H(Y|XE)_{\mathcal{C}}, \tag{8.7}$$

where the set $\mathfrak{C}_1^P[\omega,\Theta] \subset \Gamma[\omega,\Theta]$ represents the set of all pure state and projective strategies that do not allow for any pre-shared randomness - i.e. the collection of the strategies in the set $\Gamma[\omega,\Theta]$, for which the classical register Λ is trivial.

Proof. Let $\mathcal{C}$ be any strategy. Then the CQ state $\rho_{\mathcal{C}}$ is of the form

$$\rho_{\mathcal{C}} = \sum_{\lambda=1}^{n}\sum_{x,y} p(\lambda)p_X(x)\,|x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \mathrm{tr}_A\left(M_y^\lambda \otimes \mathbb{1}_E \rho_{A\tilde{E}}^{x,\lambda}\right) \otimes |\lambda\rangle\langle\lambda|_\Lambda\,.$$

[1] Note any operator P satisfying $P^2 = P$ is called a projector. Importantly, this also includes the operators $\mathbb{1}$ and zero operator O for our case.

The above motivates us to define strategies $\mathcal{C}^\lambda = (\rho_{AE}^{0,\lambda}, \rho_{AE}^{1,\lambda}, M_0^\lambda)$. By construction, every strategy $\mathcal{C}^\lambda$, is in the set $\mathfrak{C}_1[\omega_\lambda, \Theta_\lambda]$ for some ω_λ and $\Theta_\lambda \in [0,1]$.
It is easy now to see that

$$\begin{aligned}\mathcal{O}_{\Pi_0}(\mathcal{C}) &= \frac{1}{2}\sum_x \operatorname{tr}\left(\Pi_0 \otimes \mathbb{1}_E \left(p(\lambda)\rho_{A\tilde{E}}^{x,\lambda} \otimes |\lambda\rangle\langle\lambda|\right)\right) \\ &= \frac{1}{2}\sum_x p(\lambda)\operatorname{tr}\left(\Pi_0 \rho_A^{x,\lambda}\right) \\ &= \sum_\lambda p(\lambda)\mathcal{O}_{\Pi_0}(\mathcal{C}_\lambda). \end{aligned} \tag{8.8}$$

Similarly, we can show that $S(\mathcal{C}) = \sum_\lambda p(\lambda) S(\mathcal{C}_\lambda)$.
Finally,

$$\begin{aligned} H(Y|X\Lambda E)_{\mathcal{C}} &= \sum_\lambda p(\lambda) H(Y|X, \Lambda = \lambda, E)_{\mathcal{C}} \\ &= \sum_\lambda p(\lambda) H(Y|XE)_{\mathcal{C}_\lambda}. \end{aligned}$$

Combining all the above gives the objective function

$$H(Y|X\Lambda\tilde{E}) = \sum_\lambda p(\lambda) H(Y|X\tilde{E})_{\mathcal{C}_\lambda}, \tag{8.9}$$

and the constraints are given by

$$\sum_\lambda p(\lambda)\mathcal{O}_{\Pi_0}(\mathcal{C}_\lambda) = \Theta \tag{8.10}$$

$$\sum_\lambda p(\lambda) S(\mathcal{C}_\lambda) = \omega. \tag{8.11}$$

Thus, solving the for the function

$$\begin{aligned} G_{p_X}(\omega, \Theta) := \quad & \inf H(Y|X\tilde{E})_{\mathcal{C}_\lambda} \\ & \mathcal{O}_{\Pi_0}(\mathcal{C}_\lambda) = \Theta \\ & S(\mathcal{C}_\lambda) = \omega, \end{aligned} \tag{8.12}$$

and taking the convex lower bound of the function, will give the function $F_{p_X}(\omega, \Theta)$. □

In section 8.10, we have examined the methods used to calculate the lower bounds over arbitrary probability distributions $p(\mu)$. The fundamental idea of the Lemma above is that when computing rates, it is possible to restrict the calculation to a scenario in which no prior randomness has been shared, by deferring the consideration of lower bounds across all convex combinations at this stage.

8.3 REDUCTION TO QUBIT STRATEGIES

The main issue with the optimization problem $G_{p_X}(\omega, \Theta)$ is that no assumption has been made on the dimensions of the state ρ^x_{AE}. We saw in Chapter 4 that the complexity of these optimization problems is reduced using Jordan's Lemma. Jordan's Lemma shall be useful for our protocols here as well. We had stated Jordan's Lemma in Chapter 4 (see lemma 2), however, we state it here again for convenience.

Lemma 41 (Jordan's Lemma Extended)**.** *Let $A, B \in S(\mathcal{H})$ be two projections. Then*

$$\mathcal{H} = \bigoplus_{\alpha} \mathcal{H}_\alpha \tag{8.13}$$

such that each $\mathcal{H}_\alpha$ is an invariant subspace of $\mathcal{H}$ under the action of A, B, $\mathbb{1} - A$ and $\mathbb{1} - B$. Moreover, the dimension of each subspace $\mathcal{H}_\alpha$ is at most 2.

We will now prove the following technical result:

Lemma 42. *Let P_0 and Π_0 be two projectors on the Hilbert space $\mathcal{H}$. Further let Π_0 be a rank one projector. Consider the Jordan decomposition of $\mathcal{H} = \bigoplus_\lambda \mathcal{H}_\lambda$ defined by the operators $\{P_0, \Pi_0, \mathbb{1} - P_0, \mathbb{1} - \Pi_0\}$. Among these subspaces, all spaces $\mathcal{H}_\lambda$ are contained within the null space of Π_0, except for a single subspace denoted as $\mathcal{H}_{\lambda_0}$. Furthermore, the projector onto $\mathcal{H}_{\lambda_0}$ takes the form:*

$$P_\lambda = \Pi_0 + \bar{P}, \tag{8.14}$$

where $\bar{P}$ is any projection, up to rank 1, onto the null space of Π_0

Proof. From Lemma 64, we can deduce that $[\Pi_0, P_\lambda] = 0$, indicating that P_λ and Π_0 share common eigenvectors. Likewise, $[\mathbb{1} - \Pi_0, P_\lambda] = 0$ implies that $\mathbb{1} - \Pi_0$ shares eigenvectors with P_λ. Since P_λ is a projection onto a subspace of dimension at most 2, it must take the form $P_\lambda = \alpha_\lambda \Pi_0 + \beta_\lambda \bar{P}_\lambda$, where $\bar{P}_\lambda$ is any projection onto the null space of Π_0, and $\alpha_\lambda, \beta_\lambda \in \mathbb{R}$. Considering any two subspaces, $\mathcal{H}_{\lambda_1}$ and $\mathcal{H}_{\lambda_2}$, with orthogonal supports, we observe that $\alpha_{\lambda_1}\alpha_{\lambda_2} = \alpha_{\lambda_1}$, implying that $\alpha_\lambda \in [0, 1]$ for every λ. Furthermore, due to the orthogonality of the supports of $\mathcal{H}_\lambda$, $\alpha_\lambda = 1$ can only hold for a single subspace. □

We employ Jordan's Lemma to achieve a result that significantly simplifies the problem. This result lets us effectively transform the problem into one concerning only the convex combination of qubits. In essence, the approach is to employ Jordan's Lemma to express

the states ρ^x_{AE} in a block diagonal form of size 2×2, denoted by a direct sum $\bigoplus_\lambda \rho^{x,\lambda}_{AE}$. Here $\rho^{x,\lambda}_A \in \mathcal{S}(\mathcal{H}_\lambda)$ is a qubit state. Moreover, the measurement operators $\{M_0, \mathbb{1} - M_0\}$ also act on the a projective measurement on each subspace $\mathcal{H}_\lambda$ - i.e. $M_y = \bigoplus_\lambda M^\lambda_y$. This transformation effectively simplifies our problem to an optimization problem over qubits, thereby significantly reducing its complexity. Moreover, as we shall see the lemma above can be used to show that the overlap of the strategy arises solely from a single Jordan block, implying that only one qubit block is relevant for analysis. As a result, the optimization problem can be effectively viewed as a scenario in which the source shares a single pair of qubits, rather than a state with unrestricted dimensions.

Lemma 43. $G_{p_X}(\omega, \Theta) \geq G^{(2)}_{p_X}(\omega, \Theta)$, *where*

$$\begin{aligned} G^{(2)}_{p_X}(\omega,\Theta) = \inf \quad & \sum_{x\in\{0,1\}} \eta_x p_X(x) H(Y|E)_{\rho^x} \\ s.t. \quad & \forall x \in \{0,1\} : \eta_x \in [0,1] \\ & \forall x \in \{0,1\} : \rho^x_A \in S(\mathcal{H}_2) \\ & \forall y \in \{0,1\} : M_y = |\phi\rangle\langle\phi| \ \textit{ for some } |\phi\rangle \in \mathcal{H}_2 \\ & M_0 + M_1 = \mathbb{1}_2 \\ & \rho_x = \sum_{y\in\{0,1\}} |y\rangle\langle y|_Y \otimes \operatorname{tr}\left((M_y \otimes \mathbb{1}_E)\rho^x_{AE}\right) \\ & \sum_{x\in\{0,1\}} \left(\frac{1}{2}\eta_x \operatorname{tr}_A\left(M_x \rho^x_A\right)\right) \in [\omega - \sum_x \frac{1}{2}(1-\eta_x), \omega] \\ & \sum_{x\in\{0,1\}} \frac{1}{2}\eta_x \operatorname{tr}_A\left(\Pi_0 \rho^x_A\right) = \Theta, \end{aligned} \tag{8.15}$$

where $\mathcal{H}_2$ is a two dimensional Hilbert space and $\mathbb{1}_2$ is the identity on the space.

Proof. Let $\mathcal{C} = (\rho^0_{AE}, \rho^1_{AE}, M_0) \in \mathfrak{C}^1_P$. Let P_λ be the projection onto 2 dimensional sub-space $\mathcal{H}_\lambda$, where $\mathcal{H}_\lambda$ is any subspace that is invariant under the actions of the projectors $\Pi_0, M_0, \mathbb{1} - M_0$ and $\mathbb{1} - \Pi_0$. Now for the strategy $\mathcal{C}$ consider the following:

$$\begin{aligned} \rho_{\mathcal{C}} &= \sum_{y,x} p_X(x) |y,x\rangle\langle y,x| \otimes (\rho^x_{AE} M_y \otimes \mathbb{1}_E) \\ &= \sum_{y,x} p_X(x) |y,x\rangle\langle y,x| \otimes \operatorname{tr}_A\left(\rho^x_{AE}\left(\sum_\lambda P^2_\lambda \otimes \mathbb{1}_E\right) M_y \otimes \mathbb{1}_E\right) \\ &= \sum_{y,x,\lambda} p_X(x) |y,x\rangle\langle y,x| \otimes \operatorname{tr}_A\left((P_\lambda \otimes \mathbb{1}_E \rho^x_{AE} P_\lambda \otimes \mathbb{1}_E)(P_\lambda M_y P_\lambda) \otimes \mathbb{1}_E\right). \end{aligned}$$

We introduce the quantities $\eta^\lambda_x \rho^{x,\lambda}_{AE} = P_\lambda \otimes \mathbb{1}_E \rho^x_{AE} P_\lambda \otimes \mathbb{1}_E$ and $M^\lambda_y = P_\lambda M_y P_\lambda$, where $\eta^\lambda_x \in [0,1]$ (normalization constant) and $\rho^{x,\lambda}_{AE}$ are any normalized states. Furthermore, it

is easy to verify that $P_\lambda M_y P_\lambda$ is a projector and $\rho_A^{x,\lambda}$ are qubits. Further define ρ_x^λ to be the following CQ state

$$\rho_x^\lambda = \sum_y |y\rangle\langle y|_Y \otimes |x\rangle\langle x|_X \otimes \mathrm{tr}\left(M_y \otimes \mathbb{1}_2 \rho_{AE}^{x,\lambda}\right).$$

Using the concavity of the von Neumann entropy, we have that

$$\begin{aligned} H(Y|XE)_{\mathcal{C}} &= \sum_x p_X(x) H(Y|X=x,E)_{\rho_{\mathcal{C}}} \\ &\geq \sum_x p_X(x) \sum_\lambda \eta_x^\lambda H(Y|X=x,E)_{\rho_x^\lambda} \\ &\geq \sum_x p_X(x) \eta_x^0 H(Y|X=x,E)_{\rho_x^0}. \end{aligned}$$

From Lemma 42, let $\mathcal{H}_0$ be the unique subspace which is not entirely in the nullspace of Π_0. The score function for the strategy is given by:

$$\begin{aligned} S(\mathcal{C}) &= \sum_{\lambda,x} \frac{1}{2} \mathrm{tr}_A\left(\eta_x^\lambda \rho_A^{x,\lambda} M_x^\lambda\right) \\ &= \sum_x \frac{1}{2} \mathrm{tr}_A\left(\eta_x^0 \rho_A^{x,0} M_x^0\right) + \sum_{\lambda\neq 0,x} \frac{1}{2} \mathrm{tr}_A\left(\eta_x^\lambda \rho_A^{x,\lambda} M_x^\lambda\right), \end{aligned}$$

with $\omega_x^\lambda := \mathrm{tr}\left(M_x^\lambda \rho_A^{x,\lambda}\right)$. The expression above allows us to split the contribution of the strategy into a term that depends only on the single set of qubit states $\rho_A^{x,0}$ and projective measurements M_y^0, and a term that depends on the other possible qubit states and measurements. To obtain lower bound and upper bound on the score function $S(C)$, we can set $\omega_x^\lambda \in [0,1]$. As $\sum_\lambda \eta_x^\lambda = 1$, we get the following bounds:

$$\sum_x \frac{1}{2}\left(\eta_x^0 \mathrm{tr}_A\left(\rho_A^{x,0} M_x\right)\right) \leq S(\mathcal{C}) \leq \sum_x \frac{1}{2}\left(\eta_x^0 \mathrm{tr}_A\left(\rho_A^{x,0} M_x\right) + (1-\eta_x)\right).$$

The most importantly, for the overlap conditions, we have that

$$\begin{aligned} \mathcal{O}_{\Pi_0}(\mathcal{C}) &= \sum_x \frac{1}{2}\eta_x \,\mathrm{tr}\left(\Pi_0 \rho_A^x\right) \\ &= \sum_{x,\lambda} \frac{1}{2}\eta_x^\lambda \,\mathrm{tr}\left(\Pi_0 P_\lambda \rho_{AE}^x P_\lambda\right) \\ &= \sum_{x,\lambda} \frac{1}{2}\eta_x^0 \,\mathrm{tr}\left(\Pi_0 \rho_{AE}^{x,0}\right) \\ &= \Theta. \end{aligned}$$

Now, replacing $\eta_x^0 \to \eta_x$, $\rho_{AE}^{x,0} \to \rho_{AE}^x$ and $M_y^0 \to M_y$ proves the lemma. □

8.4 CONVERTING THE OPTIMIZATION PROBLEM TO A TRADITIONAL FORM

In the previous section, we showed that it is sufficient to restrict to qubit strategies. The aim now is to simplify the optimization problem obtained in the previous section to a more traditional optimization problem that can be solved using numerical techniques. We begin by simplifying the expression of the von Neumann entropy as

$$\begin{aligned} H(Y|E) &= H(Y,E) - H(E) \\ &= H(Y) + H(E|Y) - H(E) \\ &\geq H(Y) - H(E). \end{aligned}$$

We used the chain rule for conditional von Neumann entropy, as well as the fact that $H(E|Y)_\rho \geq 0$ if $\rho = \sum_y p_Y(y) \left|y\right\rangle\!\left\langle y\right| \otimes \rho_E^y$. In fact, the inequality above becomes tight when ρ_E^y is a pure state, which is the case in our setting.
Next, we express the CQ state of our protocol when $X = x$ is observed:

$$\sum_Y p_Y^x(y) \left|y\right\rangle\!\left\langle y\right| \otimes \mathrm{tr}_A \left(\frac{M_y^x \otimes \mathbb{1}_E \rho_{AE}^x}{\mathrm{tr}_A \left(M_y^x \rho_A^x \right)} \right), \tag{8.16}$$

where $p_Y^x(y) := \mathrm{tr}_A \left(\rho_A^x M_y \right)$. By noting that $p_Y^x(y) = 1 - p_Y^x(y \oplus 1)$, we can compute the expression for $H(Y)$ as

$$H(Y) = -\sum_y p_Y^x(y) \log_2 \left(p_Y^x \right) = H_{\mathrm{bin}} \left(\mathrm{tr} \left(\rho_A^x M_0 \right) \right). \tag{8.17}$$

Similarly, we can compute $H(E)$ by noting that ρ_{AE}^x is a purification of ρ_A^x. Specifically, note that

$$\begin{aligned} \rho_E^x &= \sum_y p_Y^x(y) \, \mathrm{tr}_A \left(\frac{M_y \otimes \mathbb{1} \rho_{AE}^x}{\mathrm{tr}_A \left(\rho_A^x M_0 \right)} \right) \\ &= \mathrm{tr}_A \left((\sum_y M_y) \otimes \mathbb{1}_E \rho_{AE}^x \right) \\ &= \mathrm{tr}_{AE} \left(\rho_{AE}^x \right). \end{aligned}$$

By applying Lemma 65, we can conclude that the entropies $H(\rho_E^x)$ and $H(\rho_A^x)$ are equivalent if ρ_{AE}^x is a purification of ρ_A^x.
The discussion above leads to the following result

Lemma 44. *The optimization problem (8.15) can be expressed as*

$$\begin{aligned} G^{(2)}_{p_X}(\omega,\Theta) = \inf \quad & \sum_{x\in\{0,1\}} \eta_x p_X(x) \left(H_{\text{bin}}\left(\text{tr}\left(\rho^x_A M_0\right)\right) - H(\rho^x_A)\right) \\ s.t. \quad & \forall x \in \{0,1\} : \eta_x \in [0,1] \\ & \forall x \in \{0,1\} : \rho^x_A \in S(\mathcal{H}_2) \\ & \forall y \in \{0,1\} : M_y = |\phi\rangle\langle\phi| \ \textit{ for some } \ |\phi\rangle \in \mathcal{H}_2 \\ & M_0 + M_1 = \mathbb{1}_2 \\ & \rho_x = \sum_{y\in\{0,1\}} |y\rangle\langle y|_Y \otimes \text{tr}\left((M_y \otimes \mathbb{1}_E)\rho^x_{AE}\right) \\ & \sum_{x\in\{0,1\}} \left(\frac{1}{2}\eta_x \,\text{tr}_A\left(M_x \rho^x_A\right)\right) \in [\omega - \sum_x \frac{1}{2}(1-\eta_x), \omega] \\ & \sum_{x\in\{0,1\}} \frac{1}{2}\eta_x \,\text{tr}_A\left(\Pi_0 \rho^x_A\right) = \Theta. \end{aligned} \tag{8.18}$$

8.5 OPTIMIZATION PROBLEM IN TERMS OF BOUNDED REAL VARIABLES

The problem with Lemma 8.18 is that it is expressed in terms of the states ρ^x_A and the measurement M_y. We would like to re-express this optimization problem in terms of a standard optimization problem on a bounded domain $\mathcal{D} \subseteq \mathbb{R}^n$. To achieve this, we without any loss of generality we parameterize

$$\rho^x_A = \frac{\mathbb{1}}{2} + \sum_{i=1}^{3} \frac{a^x_i}{2}\sigma_i, \quad M_0 = |\psi_0\rangle\langle\psi_0| = \frac{\mathbb{1}}{2} + \sum_{i=1}^{3} \frac{b_i}{2}\sigma_i \text{ and } \Pi_0 = |0\rangle\langle 0| = \frac{\mathbb{1}}{2} + \frac{\sigma_1}{2}. \tag{8.19}$$

Here $\{\sigma_i\}_{i=1}^3$ are Pauli x, y, z operators. The constraints $\sum_i (a^x_i)^2 \leq 1$ and $\sum_i b_i^2 = 1$ must be satisfied, as they ensure that the parameters $\{a^x_i\}_{i=1}^3$ and $\{b_i\}_{i=1}^3$ represent a valid density operator and a projective measurement, respectively. We now have an optimization problem over 11 variables- 3 parameters for each state ρ^x_A, three parameters describing the measurement operator and 2 parameters η_0 and η_1 that give the probability of the Jordan Block $\mathcal{H}_0$ occurring. Thus, the optimization problem can be cast as an optimization problem over a compact subset of $\mathbb{R}^d$, where $d = 11$ at this stage. We will now reduce the value of d to simplify the optimization problem - i.e. get rid of some redundant variables.

Before we simplify this optimization problem any further, we introduce the following function that will help keep the expressions compact

$$\Phi(x) := H_{\text{bin}}\left(\frac{1}{2} + \frac{x}{2}\right). \tag{8.20}$$

Several properties of $\Phi(x)$ including its monotocity etc. can be easily inferred from the properties of the binary entropy H_{bin}. Some other useful properties of $\Phi(x)$ that are relevant to our proof are discussed in Appendix B.1. Now consider the following result

Lemma 45. *The optimization problem (8.18) is equivalent to the following optimization problem*

$$\begin{aligned}
\inf \quad & \sum_x \eta_x p_X(x) \left(\Phi\left(a_x \cos(\theta_x - \phi)\right) - \Phi\left(a_x\right)\right) \\
s.t. \quad & \forall x \in \{0,1\} : \eta_x, a_x \in [0,1] \\
& \forall x \in \{0,1\} : \theta_x \in [0, 2\pi] \\
& \phi \in [0, 2\pi] \\
& \sum_x \frac{1}{2}\eta_x \left(\frac{1}{2} + \frac{(-1)^x a_x \cos(\theta_x - \phi)}{2}\right) \in [\omega - \sum_x \frac{1}{2}(1-\eta_x), \omega] \\
& \sum_x \frac{1}{2}\eta_x \left(\frac{1}{2} + \frac{a_x \cos(\theta_x)}{2}\right) = \Theta.
\end{aligned} \tag{8.21}$$

Proof. First, note that the objective function and constraints in the optimization problem (8.18) are expressed solely in terms of $\text{tr}_A\left(M_0 \rho_A^x\right)$, $\text{tr}_A\left(\Pi_0 \rho_A^x\right)$, and $H(\rho_A^x)$, so it is useful to express them in terms of the Bloch vectors. Define $\mathbf{a}^x$ and $\mathbf{b}$ the Bloch vector that denotes the qubit states ρ_A^x and the projection M_0. Further notice that $\Pi_0 = |0\rangle\langle 0|$ can be equivalently expressed as the vector $\hat{\mathbf{z}}$ in this notation. Thus,

$$\begin{aligned}
\text{tr}_A\left(M_0 \rho_A^x\right) &= \frac{1}{2} + \frac{\mathbf{a}^x.\mathbf{b}}{2} \\
\text{tr}_A\left(\Pi_0 \rho_A^x\right) &= \frac{1}{2} + \frac{\mathbf{a}^x.\hat{\mathbf{z}}}{2}.
\end{aligned}$$

Without any loss of generality, we can chose a basis such that $\mathbf{z} = (1,0,0)$ and $\mathbf{b} = (\cos(\phi), \sin(\phi), 0)$. Furthermore, we write that $\mathbf{a}^x = (a_x \cos(\theta_x), a_x \sin(\theta_x), \tilde{a}_x)$. In this notation, this gives:

$$\begin{aligned}
\text{tr}_A\left(M_0 \rho_A^x\right) &= \frac{1}{2} + \frac{a_x \cos(\theta_x - \phi)}{2} \\
\text{tr}_A\left(\Pi_0 \rho_A^x\right) &= \frac{1}{2} + \frac{a_x \cos(\theta_x)}{2}.
\end{aligned}$$

The entropy $H(\rho_A^x)$ can be computed in terms of its eigenvalues $\frac{1}{2}+\frac{|\mathbf{a}^x|}{2}$ and $\frac{1}{2}-\frac{|\mathbf{a}^x|}{2}$ as

$$H(\rho_A^x) = H_{\text{bin}}\left(\frac{1}{2}+\frac{|\mathbf{a}^x|}{2}\right) = \Phi\left(|\mathbf{a}^x|\right).$$

We can use the monotonicity of $\Phi(x)$ for $x>0$, so that

$$H(\rho_A^x) = \Phi\left(\sqrt{a_x^2+\tilde{a}_x^2}\right) \leq \Phi\left(a_x\right).$$

Note that when optimizing over all strategies, the inequality is tight. Combining all the results above and performing the substitutions in (8.18), we can prove the lemma. □

8.6 EXTENDING THE DOMAIN

To compute the rates, we want to solve the optimization problem 8.21 using numerical techniques. We find that this optimization problem is resource consuming to solve directly, and therefore, we aim to lower bound this problem, which can be effectively handled numerically. By Lemma 63 (see Appendix B), we can lower-bound the objective function by noting that

$$\eta_x\left(\Phi\left(a_x\cos(\xi_x)\right)-\Phi\left(a_x\right)\right) \geq \left(\Phi\left(\eta_x a_x\cos(\xi_x)\right)-\Phi\left(\eta_x a_x\right)\right). \tag{8.22}$$

Furthermore, we can replace the constraints

$$\begin{aligned}\sum_x \frac{1}{2}\eta_x\left(\frac{1}{2}+\frac{(-1)^x a_x\cos(\theta_x-\phi)}{2}\right) &\in [\omega-\sum_x\frac{1}{2}(1-\eta_x),\omega]\\ \sum_x \frac{1}{2}\eta_x\left(\frac{1}{2}+\frac{a_x\cos(\theta_x)}{2}\right) &= \Theta\end{aligned}$$

of the optimization problem (8.21) by the relaxed constraints

$$\begin{aligned}\sum_x \frac{1}{2}\eta_x\left(\frac{1}{2}+\frac{(-1)^x a_x\cos(\theta_x-\phi)}{2}\right) &\geq \omega-\sum_x\frac{1}{2}(1-\eta_x)\\ \sum_x \frac{1}{2}\eta_x\left(\frac{1}{2}+\frac{a_x\cos(\theta_x)}{2}\right) &\geq \Theta.\end{aligned}$$

Such a relaxation can only decrease the minimal value, as the feasible set of (8.21) is a subset of the relaxed feasible set. As, we shall eventually take the convex envelope of the function later on, this relaxation, does not affect the overall tightness of our rate.

It can now be useful to re-label the variables $\eta_x a_x = \tilde{a}_x$. As $a_x \in [0,1]$, we can add another constraint that $\tilde{a}_x \leq \eta_x$, completely eliminating the role of a_x in the optimization problem. We now make the following substitutions $\theta_x - \phi \to \xi_x$ Summarizing all the above leads to the following result:

Lemma 46. *A lower bound on the function $G^{(2)}_{p_X}(\omega, \Theta)$ can be achieved by computing the optimization problem*

$$\begin{aligned} \inf \quad & \sum_x p_X(x) \left(\Phi\left(\tilde{a}_x \cos(\xi_x)\right) - \Phi\left(\tilde{a}_x\right)\right) \\ s.t. \quad & \forall x \in \{0,1\} : 1 \geq \eta_x \geq \tilde{a}_x \geq 0 \\ & \forall x : \xi_x \in [0,1] \\ & \sum_x \left(-\eta_x + (-1)^x \tilde{a}_x \cos(\xi_x)\right) \geq 4\omega - 4 \\ & \sum_x \left(\eta_x + \tilde{a}_x \cos(\xi_x + \phi)\right) \geq 4\Theta. \end{aligned} \tag{8.23}$$

8.7 ELIMINATING ONE MORE VARIABLE

Now, we can immediately simplify the final constraint in terms of overlap by observing that ϕ only appears one of the constraints. We should, without any loss of generality, choose the value of ϕ that optimizes the constraint (See lemma 66 in appendix B). We observe that:

$$\begin{aligned} \sum_x \eta_x a_x \cos(\xi_x + \phi) &= \left(\sum_x \eta_x a_x \cos(\xi_x)\right) \cos(\phi) - \left(\sum_x \eta_x a_x \sin(\xi_x)\right) \sin(\phi) \\ &\leq \sqrt{\left(\sum_x \eta_x a_x \cos(\xi_x)\right)^2 + \left(\sum_x \eta_x a_x \sin(\xi_x)\right)^2}. \end{aligned} \tag{8.24}$$

Note[2] that because $\Theta - \sum_x \eta_x \geq 0$ for $\Theta \geq \frac{1}{2}$ and

$$\sqrt{\left(\sum_x \eta_x a_x \cos(\xi_x)\right)^2 + \left(\sum_x \eta_x a_x \sin(\xi_x)\right)^2} + \sum_x \eta_x \geq 4\Theta \tag{8.25}$$

implies that

$$\left(\sum_x \eta_x a_x \cos(\xi_x)\right)^2 + \left(\sum_x \eta_x a_x \sin(\xi_x)\right)^2 \geq \left(4\Theta - \sum_x \eta_x\right)^2. \tag{8.26}$$

[2] It is easy to see that $G_{p_X}(\omega, \Theta) = 0$ when $\Theta \leq \frac{1}{2}$, so $\Theta \geq \frac{1}{2}$ is the domain of our interest to begin with.

Summarizing all the above long with the results in Lemma 46 leads to the following result:

Lemma 47. *The function $\mathcal{G}_{p_X}(\omega, \Theta)$ defined by*

$$\begin{aligned} \inf \quad & \sum_x p_X(x) \left(\Phi\left(\tilde{a}_x \cos(\xi_x)\right) - \Phi\left(\tilde{a}_x\right)\right) \\ s.t. \quad & \sum_x \left(-\eta_x + (-1)^x \tilde{a}_x \cos(\xi_x)\right)\right) \geq 4\omega - 4 \\ & \left(\sum_x \tilde{a}_x \cos(\xi_x)\right)^2 + \left(\sum_x \tilde{a}_x \sin(\xi_x)\right)^2 - \left(4\Theta - \sum_x \eta_x\right)^2 \geq 0 \\ & \forall x \in \{0,1\} : 1 \geq \eta_x \geq \tilde{a}_x \geq 0. \end{aligned} \tag{8.27}$$

is a lower bound on the function $G^{(2)}_{p_X}(\omega, \Theta)$.

In the appendix B, we shall show that the function $\mathcal{G}_{p_X}(\omega, \Theta)$is monotonically increasing in ω and Θ.

8.8 COMPUTING THE OPTIMIZATION PROBLEM OVER GRIDS

We now address the problem of computing the convex envelope of the function $\mathcal{G}_{p_X}(\omega, \Theta)$. As the functional form of this is difficult to obtain, we can try to compute this numerically using known optimization techniques. However, in order to compute the convex envelope numerically, we can only compute this on a finite set of points or a finite grid. So, we construct another a lower bound $\mathcal{G}^{\mathcal{P}}_{p_X}(\omega, \Theta)$ of the function $\mathcal{G}_{p_X}(\omega, \Theta)$, that can be only computed using a finite collection of points. For $\Theta > \frac{1}{2}$ define:

$$\mathcal{G}^{\mathcal{P}}_{p_X}(\omega, \Theta) := \mathcal{G}_{p_X}(\omega_{\mathcal{P}}, \Theta_{\mathcal{P}}), \tag{8.28}$$

where

$$\begin{aligned} \omega_{\mathcal{P}} &:= \max\{\omega_i : (\omega_i, \Theta_j) \in \mathcal{P}, \omega_i \leq \omega, \Theta_j \leq \Theta\} \\ \Theta_{\mathcal{P}} &:= \max\{\Theta_j : (\omega_i, \Theta_j) \in \mathcal{P}, \omega_i \leq \omega, \Theta_j \leq \Theta\}. \end{aligned}$$

Note that $\mathcal{G}^{\mathcal{P}}_{p_X}(\omega, \Theta)$ is a lower bound on the function $\mathcal{G}_{p_X}(\omega_{\mathcal{P}}, \Theta_{\mathcal{P}})$ as $\mathcal{G}_{p_X}(\omega_{\mathcal{P}}, \Theta_{\mathcal{P}})$ is also monotonically increasing in ω and Θ; precisely, $\partial_\omega \mathcal{G}_{p_X} \geq 0$ and $\partial_\Theta \mathcal{G}_{p_X} \geq 0$.
Using the algorithms described in the section 8.10, the function $\mathcal{F}_{p_X}(\omega, \Theta) := \text{convenv}(\mathcal{G}^{\mathcal{P}}_{p_X}(\omega, \Theta))$ can be computed in linear time over the partition $\mathcal{P}$, which can be extended to the entire domain using the same trick above.

Note that $\mathcal{F}^{\mathcal{P}}_{p_X}$ can be shown to be a lower bound of the F_{p_X} in the section 5.2.4. The optimization problem for $\mathcal{G}^{\mathcal{P}}_{p_X}(\omega, \Theta)$ can be easily relaxed to a polynomial optimization problem by lower bounding the objective function

$$\Phi(\tilde{a}_x \cos(\xi_x)) - \Phi(x) \to \Phi_n(\tilde{a}_x \cos(\xi_x)) - \Phi_m(\tilde{a}_x) - I_{m+1},$$

and by replacing the variables $\cos(\xi_x) \mapsto c_x, \sin(\xi_x) \mapsto d_x$ with an additional constraint $c_x^2 + d_x^2 = 1$. Here Φ_n, Φ_m are appropriate polynomial approximations of the function $\Phi(x)$ and $I_{m+1} := \sup_{x\in[-1,1]} |\Phi(x) - \Phi_m(x)|$. In the next section, we shall explicitly construct such polynomial approximations.

8.9 CONVERTING THE OPTIMIZATION PROBLEM TO A POLYNOMIAL OPTIMIZATION PROBLEM

In this section, we approximate the binary entropy function (and its difference) in terms of polynomials.

We begin by the integral representation of the binary entropy

$$\log(x) = \frac{1}{\ln(2)} \int_0^1 \frac{x-1}{t(x-1)+1} \mathrm{d}t. \tag{8.29}$$

From this representation, we can derive the integral representation of the function $\Phi(x) := H_{\mathrm{bin}}(\frac{1}{2} + \frac{x}{2})$ as

$$\Phi(x) = \int_0^1 (1-x^2) \frac{(2-t)}{(2-t(1-x))(2-t(1+x))} \mathrm{d}t. \tag{8.30}$$

Upon performing some rearrangements, and performing change in variable, we obtain the following expression

$$\Phi(x) = \int_{\frac{1}{2}}^1 \frac{1}{z\ln(2)} \left(\frac{1-x^2}{1 - \left(\frac{1-z}{z}\right)^2 x^2} \right) \mathrm{d}z \tag{8.31}$$

Now as $\frac{1-z}{z} x \in [0,1]$ in the range $z \in [\frac{1}{2}, 1]$ and $x \in [-1,1]$, we can expand the integrand to the following convergent infinite sum

$$\Phi(z) = \sum_{n=0}^{\infty} \left(\int_{\frac{1}{2}}^1 \frac{1}{z\ln(2)} \left(\frac{1-z}{z}\right)^{2n} \mathrm{d}z \right) x^{2n}(1-x^2). \tag{8.32}$$

Note here that the integrals

$$I_n := \int_{\frac{1}{2}}^{1} \frac{1}{z\ln(2)} \left(\frac{1-z}{z}\right)^{2n} \mathrm{d}z. \tag{8.33}$$

are analytically solvable, e.g. $I_0 = 1$, $I_1 = 1 - 1/(2\ln(2))$, $I_3 = 1 - 7/(12\ln(2))$ and so on. Furthermore, using this technique, we obtain lower-bounds on the binary entropy as $I_n(x) \geq 0$ can be trivially established. An approximation for the binary entropy can be made by truncating the summation to a certain term

$$\Phi(x) \approx \Phi_n(x) := \sum_{k=1}^{n} I_k x^{2k}(1-x^2). \tag{8.34}$$

The error term is given by

$$\begin{aligned} \Phi(x) - \Phi_n(x) &= (1-x^2) \sum_{k=n+1}^{\infty} I_k x^{2k} \\ &\leq (1-x^2) I_{n+1} \sum_{k=n+1}^{\infty} x^{2k} \\ &= I_{n+1} x^{2(n+1)}, \end{aligned}$$

where we have used the fact that $I_n > I_{n+k}$ for every $k \in \mathbb{N}$. We lower bound the objective function in (8.23) by introducing the functions

$$P_n(\tilde{a}, \cos(\xi)) := \Phi_n(\tilde{a}\cos(\xi)) - \Phi_n(\tilde{a}) - I_{n+1}. \tag{8.35}$$

There is still some work to do in order to lower bound the optimization problem (8.23) to a polynomial problem, as $\cos(\xi_x)$ are not polynomials. However, we can identify the functions $\lambda_{1,x} = \tilde{a}_x \cos(\xi_x)$ and $\lambda_{2,x} = \tilde{a}_x \sin(\xi_x)$. Now, if treat $\lambda_{i,x}$ as free variables in our optimization problem by introducing additional constraints

$$\sum_i \lambda_{i,x}^2 = \tilde{a}_x^2,$$

then we can lower bound (8.23) to a polynomial optimization problem. Lower bounds can now be obtained to $\mathcal{P}_1(\Theta, \omega)$ by solving this polynomial optimization problem using a sum-of-squares relaxation by using software such as Ncpol2sdpa. The final optimization

problem is now of the form

$$\begin{aligned}\mathcal{P}_1(\omega,\Theta) := \inf \quad & \sum_x p_X(x)\left(\Phi_n\left(\lambda_{1,x}\right)\right) - \Phi_n\left(\tilde{a}_x\right)) - I_{n+1} \\ \text{s.t.} \quad & \sum_x \left(-\eta_x + (-1)^x \lambda_{1,x}\right)) \geq 4\omega - 4 \\ & \left(\sum_x \lambda_{1,x}\right)^2 + \left(\sum_x \lambda_{2,x}\right)^2 \geq \left(4\Theta - \sum_x \eta_x\right)^2 \\ & \lambda_{1,x}^2 + \lambda_{2,x}^2 = \tilde{a}_x \\ & \eta_x \geq \tilde{a}_x \\ & 1 \geq \eta_x \\ & \tilde{a}_x \geq 0.\end{aligned} \tag{8.36}$$

Note that this is a constrained polynomial optimization problem over 8 real variables. Actually, the number of variables is 6, and we have two dummy variables here. Thus, we can summarize the final result of the chapter:

Theorem 4. *Let $\mathcal{P}_1(\omega,\Theta)$ be defined as in (8.36), then*

$$\inf_{\mathcal{C}\in\Gamma[\omega,\Theta]} H(Y|XE)_{\rho_{\mathcal{C}}} \geq \textit{convenv}\left(\mathcal{P}_1(\omega,\Theta)\right). \tag{8.37}$$

8.10 TAKING THE CONVEX LOWER-BOUND OF THE RATE

Upon computing the rate for a single strategy, now we are in the position to compute the rate for the convex combination of such strategies. In this section, we shall discuss a method that allows us to compute the convex envelope of any function $f : \mathbb{R}^n \mapsto \mathbb{R}$. We start by defining the following

Definition 21 (Convex envelope)**.** Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be any function, then the convex envelope of f given by convenv(f) is the function

$$\text{convenv}(f) := \max\{g : g \text{ is convex and } \forall \mathbf{x} :\ g(\mathbf{x}) \leq f(\mathbf{x})\}. \tag{8.38}$$

Here $\max$ is taken to be point-wise maximum.

In simple words, convenv(f) is the biggest convex function that is not greater than f. In other words, the convenv(f) is the solution of the optimization problem

$$\begin{aligned}\text{convenv}(f(\mathbf{x}^*)) = \inf_{\mu} \quad & \int \mathrm{d}\mu(\mathbf{x}) f(\mathbf{x}_\mu) \\ \text{s.t.} \quad & \int \mathrm{d}\mu(\mathbf{x})\mathbf{x} = \mathbf{x}^*,\end{aligned} \tag{8.39}$$

where $\inf$ is taken over all the probability measures. We now define an important tool in convex analysis that allows us to compute the convex envelope of a function.

Definition 22 (Legendre-Fenchel transform)**.** Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be any function. Then $f^* : \mathbb{R}^n \mapsto \mathbb{R}$ is its Legendre-Fenchel transform if

$$f^*(\mathbf{k}) := \sup_{\mathbf{x}\in\mathbb{R}^n} \left(\langle k, x\rangle - f(\mathbf{x})\right). \tag{8.40}$$

The following can be shown for $f^*(\mathbf{x})$

Lemma 48. *Let $f : \mathcal{D} \mapsto \mathbb{R}$ be bounded. Then the following holds*

- *f^* is convex*
- $(f^*)^* = \text{convenv}(f)$

The proof of the lemma above can be found in textbooks of convex analysis such as [88] and shall be omitted in the work. There are algorithms in available in the literature [89, 90] to compute the convex envelopes by computing the LF conjugate of the function twice. We compute the convex envelope using the method of [90], the code for which was generously provided by the authors to us.
Finally, we prove the final result about the fact that rate is non-decreasing.

Lemma 49. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a convex function such that $f \geq 0$. Suppose that $f(x, y) = 0$ for $x \leq x_0$ or $y \leq y_0$ for some $x_0, y_0 \in \mathbb{R}$. Then $f(x, y)$ is non-decreasing.*

Proof. Since f is convex, the functions $g_y := f(\cdot, y)$ and $h_x := f(x, \cdot)$ are convex as well.
Take $x' > x$ and $\mu \in [0, 1]$ such that $\mu x' + (1 - \mu)x_0 = x$. Then, by the convexity of g_y, we have

$$\mu g_y(x) + (1 - \mu)g_y(x') \geq g_y(\mu x' + (1 - \mu)x_0) = g_y(x),$$

which implies $\mu f(x', y) \geq f(x, y)$ for all y. Thus, we have $f(x', y) \geq f(x, y)$ for all y, showing that f is non-decreasing in x. Similarly, we can show that f is non-decreasing in y using the convexity of h_x.
Therefore, $f(x, y)$ is non-decreasing in both variables. □

— 9 —

Results and discussion

In the previous chapter, we discussed how to find the minimum of the single round von Neumann entropies for strategies that achieve a particular score and have a threshold overlap. In this chapter, we will look into some semi-DI protocols and apply our results to compute randomness rates for these protocols.

In protocols of randomness expansion, we assume that the input randomness is not accessible to any eavesdropper during each round of the protocol. This assumption is crucial for the security of the protocol, as if the Eavesdropper knows the input randomness, then they can easily pre-program the devices to produce deterministic outputs. We design two types of protocols, depending on the source of input randomness: private and public. Before discussing the protocols in detail, we have a discussion regarding the assumptions under which our protocol operates.

Beyond the assumption that the source and the measurement device do not share any pre-existing entanglement, and that we have access to a trusted power meter, we have not made any assumptions regarding the functioning of the source and the measurement device. It remains an open question whether having this pre-existing entanglement offers any significant advantage to an eavesdropper's ability to predict the inputs and outputs of the protocol.

However, it is worth noting that we assume the source and the measurement device can only communicate through the signal ρ^x sent by the source and cannot communicate by any other means. In order to ensure this, similar to the DI case, the source and measurement devices must be properly shielded. Much like in Device Independent scenarios, once the protocol commences, neither the source nor the measurement device can convey information to potential adversaries. After the protocol, these devices must remain isolated (see Appendix A.5) and should not be used again.

In the traditional prepare-and-measure scenario, it is crucial to recognize that the source and the measurement device might access alternative communication avenues. These methods could potentially convey the input's information indirectly, rather than encoding it directly within the source. Consider a scenario where the source sends signals ρ^x with time lags, Δt, based on input X. This makes it straightforward for the measurement device to identify the input and send outputs using a pre-determined strategy. Ideally, these attacks can be avoided by introducing an additional abort condition which is based on the statistics of the test round of the protocol. In other words, our protocol should abort if the power meter, our trusted component, detects any anomaly in signal reception timings (for instance, the protocol aborts if the root mean square value of the time lag between consecutive signals is bigger than a threshold). However, this study does not address preventing such attacks.

A practical method to circumvent both timing and entanglement-based attacks in semi-Device Independent protocols is to acquire the source and measurement devices from separate manufacturers. While this strategy does not guarantee absolute protection, it is a reasonable precaution in real-world scenarios, based on the security demands of the protocols.

For the security of the protocol we use a composable security definition, as detailed in Chapter 3. In the next two sections, we discuss two semi-Device Independent protocols based on the setup described in Section 7.3. These protocols are inspired by the protocols 2 and 3 in Chapter 6.

9.1 RECYCLED INPUTS PROTOCOL

As discussed in the Chapter 6, in a standard randomness expansion protocol, it is assumed that the initial randomness is a private source of randomness. If this is the case, there is no incentive to make this initial randomness public. In fact, this input randomness should be recycled, and run through the randomness extractor along with the string of bits generated via the measurement device. This idea forms the basis for the following protocol:

Protocol 4. Parameters:
n – number of rounds
p_0 – probability of 1 for random number generator R_A (taken to be below $1/2$)
γ – probability of 1 for random number generator R_γ (taken to be below $1/2$)

ω_{exp} – expected score.
Θ_{exp} – expected overlap.
δ_Θ – confidence width for the overlap
δ_ω – confidence with for the score

1. Set $i = 1$ for the first round, or increase i by 1.
2. Use R_A to choose $X_i \in \{0, 1\}$, which is input to the device S. Here $X_i = 1$ occurs with probability p_0. The device S prepares a state ρ_i^x (unknown) and sends it to BS.
3. Use R_T to choose $T_i \in \{0, 1\}$, which is input to BS. BS sends the state to M if $T_i = 0$ or sends it to PM if $T_i = 1$.
4. If $T_i = 0$ (Generate round): M receives ρ_i^x and outputs $Y_i \in \{0, 1\}$. Set $U_i = (T_i, X_i, Y_i)$.
5. If $T_i = 1$ (Test round): PM receives ρ_i^x and outputs Y_i. Set $U_i = (T_i, X_i, Y_i)$.
6. Return to Step 1 unless $i = n$.
7. Compute the empirical scores $U_\#$ and $\omega_\#$ as
$$\begin{aligned} U_\# &:= \frac{1}{2}\sum_i \frac{|\{i : U_i = (0, x_i, 1)\}|}{np_X(x)\gamma}, \\ \omega_\# &:= \frac{1}{2}\sum_i \frac{|\{i : U_i = (1, x_i, x_i)\}|}{np_X(x)(1-\gamma)}. \end{aligned}$$
8. Abort the protocol if either of the conditions are not met
   a) $\omega_\# > \omega_{\text{exp}} - \delta_\omega$.
   b) $U_\# > \Theta_{\text{exp}} - \delta_\Theta$.
9. Process the concatenation of all the outputs with a quantum-proof strong extractor Ext to yield $\mathrm{Ext}(\mathbf{XY}, \mathbf{R})$, where $\mathbf{R}$ is a random seed for the extractor. Since a strong extractor is used, the final outcome can be taken to be the concatenation of $\mathbf{R}$ and $\mathrm{Ext}(\mathbf{XY}, \mathbf{R})$.

9.2 A PROTOCOL TO GENERATE PRIVATE RANDOMNESS

There maybe situations in which the initial source of randomness comes from a public source of randomness such as a randomness beacon. Such a source can also be used for

randomness expansion as long as the devices are prepared and put in a secure lab before the public randomness is made accessible. In such a scenario, the input randomness cannot be recycled, as any adversary will have access to the input randomness. Nonetheless, the public source of randomness can be converted to a source of private randomness. The only difference in this protocol would be to replace the Step 9 to Step 9' in the previous protocol (Protocol 4), where instead of extracting the key from $\mathbf{XY}$, we only extract the key from $\mathbf{X}$.

Protocol 5. Parameters:
n – number of rounds
p_0 – probability of 1 for random number generator R_A (taken to be below $1/2$)
γ – probability of 1 for random number generator R_γ (taken to be below $1/2$)
ω_{exp} – expected score.
Θ_{exp} – expected overlap.
δ_Θ – confidence width for the overlap
δ_ω – confidence with for the score

1. Set $i = 1$ for the first round, or increase i by 1.
2. Use R_A to choose $X_i \in \{0, 1\}$, which is input to the device S. Here $X_i = 1$ occurs with probability p_0. The device S prepares a state ρ_i^x (unknown) and sends it to BS.
3. Use R_T to choose $T_i \in \{0, 1\}$, which is input to BS. BS sends the state to M if $T_i = 0$ or sends it to PM if $T_i = 1$.
4. If $T_i = 0$ (Generate round): M receives ρ_i^x and outputs $Y_i \in \{0, 1\}$. Set $U_i = (T_i, X_i, Y_i)$.
5. If $T_i = 1$ (Test round): PM receives ρ_i^x and outputs Y_i. Set $U_i = (T_i, X_i, Y_i)$.
6. Return to Step 1 unless $i = n$.
7. Compute the empirical scores $U_\#$ and $\omega_\#$ as

$$\begin{aligned} U_\# &:= \frac{1}{2}\sum_i \frac{|\{i : U_i = (0, x_i, 1)\}|}{np_X(x)\gamma} \\ \omega_\# &:= \frac{1}{2}\sum_i \frac{|\{i : U_i = (1, x_i, x_i)\}|}{np_X(x)(1-\gamma)} \end{aligned}$$

8. Abort the protocol if either of the conditions are not met

a) $\omega_{\#} > \omega_{\text{exp}} - \delta_{\omega}$.

b) $U_{\#} > \Theta_{\text{exp}} - \delta_{\Theta}$.

9. Process the concatenation of all the outputs with a quantum-proof strong extractor Ext to yield $\mathrm{Ext}(\mathbf{X}, \mathbf{R})$, where $\mathbf{R}$ is a random seed for the extractor. Since a strong extractor is used, the final outcome can be taken to be the concatenation of $\mathbf{R}$ and $\mathrm{Ext}(\mathbf{X}, \mathbf{R})$.

9.3 RATES OF THE PROTOCOL

We will now discuss the computation of rates for the protocol. Similar to the protocols for DIRNE, we can use the Entropy Accumulation Theorem (EAT) to calculate the rates [14, 32]. However, during the course of our project, a more robust extension of EAT called the generalized Entropy Accumulation Theorem was published [30], which can also be used to compute the finite round key rates. It should be noted that the computation of finite round rates for the protocol is not covered in this thesis and is left as future work.

Roughly speaking, according to the EAT the asymptotic rates for our protocol can be computed by computing the infimum of the single round conditional von Neumann entropy $H(Y|XE)$ over all possible strategies that achieve a score ω and an overlap Θ. Given that the protocol consists of two distinct round types, the channel describing a single round of a protocol $\mathcal{N}$ can be expressed as

$$\mathcal{N} = (1-\gamma)\mathcal{N}_G + \gamma\mathcal{N}_T,$$

where $\mathcal{N}_G$ and $\mathcal{N}_T$ are the EAT channels corresponding to the generate and the test rounds and γ is the testing probability.

The EAT channel corresponding to the generate round is given by

$$\mathcal{N}_G : \rho_{AE} \mapsto \rho_{\mathcal{C}},$$

where ρ_{AE} is some quantum state initial quantum state of the form 8.4 and $\rho_{\mathcal{C}}$ is the state of the form 8.6. The channel describing the test round is of the form:

$$\mathcal{N}_T(\rho_{AE}) = \sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \mathrm{tr}_A\left(\Pi_0 \rho_x^{\lambda}\right) \otimes |\lambda\rangle\langle\lambda|_{\Lambda}\,. \tag{9.1}$$

We can use the concavity of the conditional von Neumann entropy to obtain a lower bound

$$H(Y|XE)_{\mathcal{N}(\rho_{AE})} \geq (1-\gamma)H(Y|XE)_{\mathcal{N}_G(\rho_{AE})}.$$

In other words, the output randomness in a protocol rand_{out} can be computed using the conditional von Neumann entropy from the generate rounds only. This justifies our choice of optimizing $H(Y|XE)_{\rho_{\mathcal{C}}}$ in the previous chapter.

Now, let us focus our discussion on the asymptotic rates for individual protocols. In protocol 5, the randomness expansion per unit round can be computed by

$$\text{rand}_{\text{out}} - \text{rand}_{\text{in}} = (1-\gamma)H(Y|XE) - H(X), \tag{9.2}$$

where $H(X)$ denotes the input randomness, which can be computed to be $H_{\text{bin}}(p_X(0))$. Note that the entropy $H(Y|XE)$ here is the entropy for generate rounds only.

For Protocol 4, since both the input and output strings are used, the difference in the output randomness (in the asymptotic limit) is given by

$$\text{rand}_{\text{out}} - \text{rand}_{\text{in}} = (1-\gamma)H(XY|E) - H(X) = (1-\gamma)H(Y|XE) - \gamma H(X), \tag{9.3}$$

where we have used the chain rule for the conditional von Neumann entropy. Again, here, the entropy $H(Y|XE)$ is the entropy of generate rounds only. Thus, in order to compute the rates of the protocols, one should compute the quantity

$$F_{p_X}(\omega, \Theta) := \inf H(Y|XE),$$

where the infimum is taken over all the single round strategies that achieve a score ω in the generate rounds and have an overlap Θ.

9.4 RESULTS

In the previous chapter (Chapter 8), we calculated the asymptotic rates $F_{p_X}(\omega, \Theta)$. The rates are shown in Figure 9.2 as a function of ω for various overlap values.

As discussed in Chapter 8, the function $F_{p_X}(\omega, \Theta)$ can be obtained by computing the convex envelope (i.e., the smallest convex function below G) of $G_{p_X}(\omega, \Theta)$, which is derived from qubit strategies. Importantly, not all values of the tuple $(\omega, \Theta) \in [1/2, 1] \times [1/2, 1]$ are attainable using a quantum strategy (see [9, 69]). In such situations, our optimization problem for $G_{p_X}(\omega, \Theta)$ also does not show any feasible solutions. To

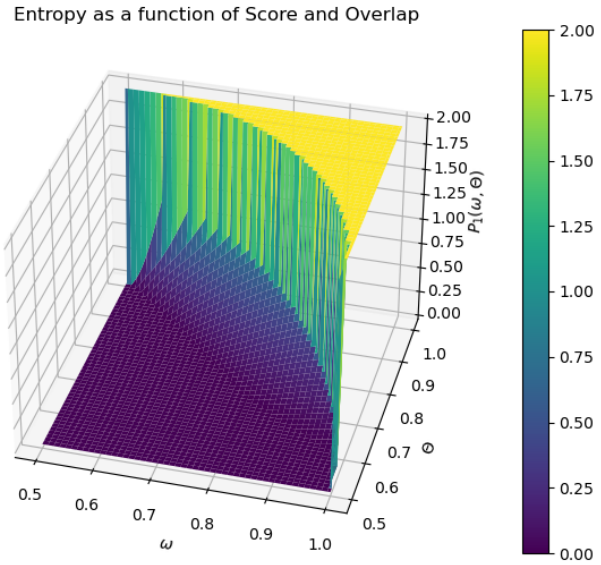


Figure 9.1: Lower bound of the function G over the extended domain $(\omega, \Theta) \in [1/2, 1] \times [1/2, 1]$ in the case when $p_X(0) = \frac{1}{2}$. The region with $G(\omega, \Theta) = 2$, is the region where no quantum strategies are found.

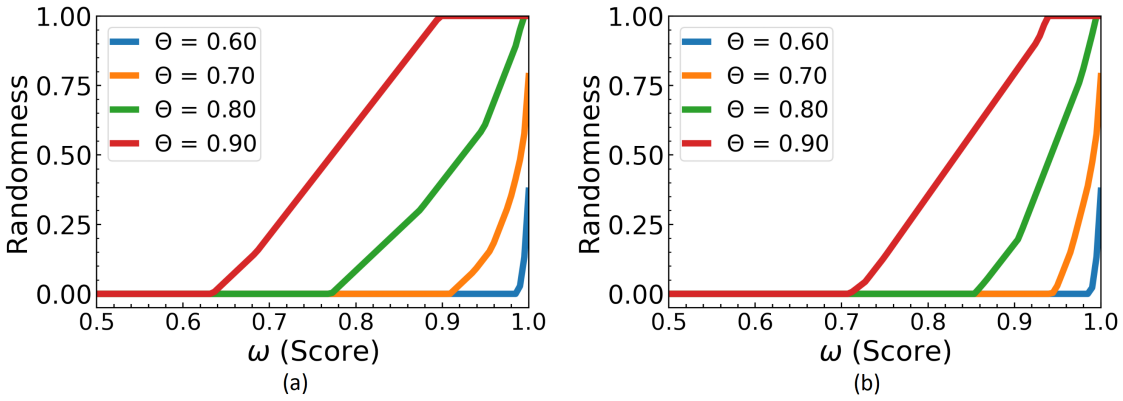


Figure 9.2: Asymptotic rates in the generate round of our protocol. Figure (a) are the rates when $p_X(0) = \frac{1}{2}$ and Figure (b) are the rates when $p_X(0) = \frac{1}{100}$. Note that the rates reported are valid in the limit $\gamma \to 0$.

determine the convex envelope of $G_{p_X}(\omega, \Theta)$ using numerical algorithms (see [89, 90] for fast algorithms for convex envelope), we extend the domain of $G_{p_X}(\omega, \Theta)$ to the entire range of $[1/2, 1] \times [1/2, 1]$ and assign it an arbitrarily high value when no feasible

quantum strategy is found. This extension results in a function $F_{p_X}(\omega, \Theta)$ over the domain $[1/2, 1] \times [1/2, 1]$. The function reported in Fig 9.2 as $F_{p_X}(\omega, \Theta)$ is a lower bound on the randomness rate if there is a quantum achievable strategy. For the regions of the extended domain where no quantum achievable strategy exists, the values of $F_{p_X}(\omega, \Theta)$ should be disregarded. Nonetheless, this extended domain proves useful when EAT is used to get bounds on finite round rates. In particular, it turns out that the min-tradeoff function needs to be defined on the extended domain,.

As anticipated, the protocol yields a high randomness rate for a fixed score (assuming quantum strategies can achieve this) when the overlap is significant. Conversely, with minimal overlap, a higher score is essential for generating randomness. This stems from the fact that deterministic strategies, akin to local deterministic strategies in the DI case, can only attain a low score. For instance, achieving a score around 0.5 is feasible through mere chance combined with some pre-programming of the devices. To get a large score, a genuine quantum strategy is required, ensuring the presence of randomness in the outputs.

For quantum achievable strategies, our results are promising. For instance, we identify the strategy $\Theta = 0.8$ and $\omega = 0.878$ as attainable using quantum theory, achieving approximately 0.319 bits per round when $p_X(0) = \frac{1}{2}$. As shown in the figure, a higher overlap yields better values of randomness rounds. However, it is crucial to consider that the power meter used in the experiment will have a detector efficiency $\eta < 1$. Consequently, for security reasons, we must adopt a pessimistic approach and underestimate the experimental overlap by a factor of η. Thus, under this pessimistic scenario, an experimental setup (consisting of good power meters with high detector efficiency) would likely result in an overlap value less than 0.8. Therefore, one should anticipate a randomness expansion of approximately $0.3(1 - \gamma)$ bits per round using this protocol, where γ represents the testing probability.

Furthermore, we have plotted the function $F_{p_X}(\omega, \Theta)$ for the scenario where $p_X(0) = \frac{1}{100}$. In this case, the protocol resembles a spot-checking type protocol, where the input randomness is negligible because most of the time $X = 1$ is being sent. Only occasionally, $X = 0$ is sent as a spot check. Since this behavior is unknown to the adversary, for the protocol not to abort, the devices must function honestly during most of the rounds. As expected, this spot-checking protocol provides a lower randomness rate. However, if the figure of merit is the ratio of output to input randomness, then having $p_X(0) \ll 1$ may be more suitable. Furthermore, for a finite number of rounds, having $p_X(0) \ll 1$, similar to the spot-checking protocol, faces challenges in providing good rates. A significantly

large number of rounds is needed to gather enough statistics to be confident in ω and Θ, especially given the scarcity of rounds with $X = 0$.

9.5 DISCUSSION

In this work, we have conducted an analysis of the semi-Device Independent protocol based on the energy and overlap bounds. We have proposed a protocol that recycles input randomness for the semi-Device Independent protocols and have also introduced another protocol that converts public randomness to private randomness.

The structure and analysis of these protocols closely resemble Device Independent protocols for randomness expansion. As a result, there are numerous opportunities for applying similar techniques and ideas from Device Independence to enhance these protocols. Just like in Device Independence, alternative score functions other than the CHSH score, as discussed in references [45, 46], might yield better performance in our semi-DI protocols. However, determining the optimal functional form of the score remains an open question. Furthermore, the choice of score can significantly impact the performance of the protocols, especially given available experimental statistics.

There is a possibility of exploring connections with concepts like self-testing, as introduced in [91] that often occur in the discussions of the DI protocols. Self-tests can yield interesting score definitions that could contribute to a partial answer to the question of finding the best score given experimental statistics.

Moreover, developing an NPA-like hierarchy for such protocols [52], could be beneficial. Such a hierarchy may allow us to use results for optimizing von Neumann entropies, such as those presented in [43], to analyze these protocols for arbitrary inputs and outputs. The current proof based on Jordan's lemma is limited to protocols with only 2 inputs and 2 outputs, so a more generalized approach is desirable.

— 10 —

Conclusion and outlook

In this section of the thesis, we delved into a comprehensive exploration of randomness expansion protocols based on minimal assumptions regarding the inner workings of the devices. The thesis commenced with an exploration of the need for randomness expansion protocols. Subsequently, we delved into various protocols for achieving randomness expansion. In Chapter 2, we discussed the pivotal role of optimization problems within the context of randomness expansion protocols. We discussed the significance of semi-definite programs in solving polynomial optimization problems. Additionally, we identified the role of min-entropy as a suitable measure for quantifying randomness in a cryptographic scenario. The presentation of the Entropy Accumulation Theorem, as an approach for establishing reliable bounds on min-entropy in randomness expansion protocols, was also discussed at the end of the chapter.

Shifting focus to protocols based on the violation of (generalized) CHSH inequalities in Chapter 3, we explored the concept of Bell's theorem within the context of randomness expansion. Chapters 4 and 5 delved into strategies for establishing tight lower bounds on conditional von Neumann entropy relevant to DIQKD and DIRNE protocols which rely on violating generalized CHSH inequalities. Our work culminated in Chapter 6, where we introduced and compared three CHSH-based protocols for randomness expansion: the spot-checking protocol (with randomness extracted from both parties), a protocol that recycles input randomness, and a protocol leveraging heavily biased input randomness. Notably, the latter two protocols close the locality loophole. Upon comparative analysis, our results reveal that the protocols that recycle input randomness exhibit exponential enhancements in the finite-round regime, rendering DIRNE protocols notably more practical.

Subsequently, the exploration extended to semi-Device Independent (semi-DI) protocols

of randomness expansion. In Chapter 7, the foundational structure of our semi-DI protocol was introduced, focusing on the preparation of states with low energy or high overlap in a system having a unique ground state. The subsequent chapter, 8, dealt with the computation of lower bounds on the von Neumann entropic quantity $H(Y|XE)$ for protocols that attain a fixed a overlap with the ground state and a fixed score. Notably, in the final chapter of semi-DI protocols (Chapter 9) we introduced two protocols - one for randomness expansion and another facilitating the conversion of public randomness to private randomness.

Throughout this thesis, we presented significant enhancements in both Device Independent and semi-Device Independent randomness expansion protocols. Our contributions include the development of superior protocols in terms of rates and security. Furthermore, we devised novel approaches to establish tighter bounds on the randomness rate. An important advantage of our work lies in our ability to recycle input randomness for our protocols. The ability to recycle input randomness remarkably yields exponential increases in randomness rates, especially for finite rounds.

While our methodology proficiently computes rates for conventional score definitions, such as the CHSH score in the DI case and the guessing probability in the semi-DI scenario, it can adapt to other linear functions of input-output statistics for generating randomness. We acknowledge the potential of alternative functionals, as demonstrated in [45, 46], which might better suit randomness expansion and QKD.

An additional avenue for improvement involves harnessing the potential offered by Multi-partite Bell inequalities. Protocols based on these inequalities offer the promise of generating up to n bits of randomness. While experimentally violating these Multi-partite Bell inequalities proves to be an even greater challenge compared to the CHSH inequality, the theoretical exploration of these scenarios remains invaluable. Such investigations enable us to grasp the true extent of quantum theory's capabilities and equip us with the necessary insights for a future where their implementation becomes more feasible.

Another avenue of investigation is to for look protocols based on multiple inputs and outputs, rather than restricting to the two input-two output case. It is worth noting that our existing protocols are tailored for binary inputs and outputs due to the reliance on Jordan's Lemma. Unfortunately, no equivalent principle to Jordan's Lemma exists for scenarios involving non-binary inputs or outputs.

Although there are approaches to bypass the limitations of Jordan's Lemma, these strategies often come with resource-intensive requirements, and/or they do not give good randomness rates. In particular, a computationally feasible way is to obtain bounds

on the randomness for multi-partite scenario is using the NPA hierarchy [52] and lower bounding the von Neumann entropy in terms of the min-entropy [51]. However, this method loses tightness. Another way to obtain bounds on the randomness is by adopting the approach from the recently introduced seminal work by Brown et.al. [57]. As it stands, this approach proves challenging for practical execution in multi-partite scenarios, even in tri-partite scenarios where randomness is drawn from all the three parties. However, in simpler cases such as the two-input two-output bipartite scenarios, this method proves to be highly practical. Therefore, it might be worthwhile to look into ways in which this method can be made less resource extensive.

Interestingly, when it comes to DIRNE protocols, there isn't always a need to close all loopholes in practical implementations. The importance of addressing the locality loophole is somewhat lessened since, in the DIRNE protocols, all devices function within a single lab. Protective shielding methods, implemented to avoid any unwanted external data transmissions, can be also be used to shield the devices themselves and eliminate any potential for device-to-device communications. While obstacles linked to the detection loophole remain, it's worth noting that to exploit this loophole, an eavesdropper would need to craft near-perfect detectors and covertly install them in the devices, which itself is challenging for an adversary to achieve practically.

In this context, the fair sampling assumption [61, 92] also is essential when performing DI protocols. This assumption forms the foundation for constructing semi-DI protocols based on the violation of a Bell inequality. Consequently, this provides further rationale for developing protocols rooted in different Bell inequalities, even with present day technology.

To bridge the gap between theory and experimentation, an ideal protocol should align with the best theoretical strategy, thereby achieving optimal rates. Experimental outcomes, however, often diverge from theoretical predictions due to factors like noise, device imperfections, or possible eavesdropping. As a result, statistics gathered from experiments should then be utilized to compute a lower bound on randomness rate. This raises the question: what is the best computed lower bound given the experimental statistics obtained? At first glance, one might consider using all available statistics and employ a method like the one presented in our thesis to determine the best possible rate. Intuitively, this seems to be the optimal method, and is indeed the case in the asymptotic limit (i.e. for very large number of rounds). Nonetheless, this method has limitations in finite-round cases, primarily because the error term in the Entropy Accumulation Theorem deteriorates when the rate is determined by a very detailed abort condition of the protocol. This

translates to the scenario in which if we condition the protocol on the full experimentally achieved input-output distribution $p_{AB|XY}$, then the error term is significantly worse than if we merely condition on the observation of a CHSH score, which is only a function of the full distribution $p_{AB|XY}$.

Thus, a trade-off emerges when considering which statistics to condition upon for finitely many rounds. It is therefore crucial to strike the right balance: over-conditioning can worsen the error term while enhancing the primary term. The practicality of various conditioning strategies varies with the finite round regime, and pinpointing the best conditions often requires a case-by-case assessment. It is also worth noting that while the error term of EAT provides a lower bound, it might not be tight. For instance, the error terms of EAT were revised in [32] and subsequently in [31], improving the error term from the original EAT statement. Looking ahead, we might see even more refinements to the error term of EAT, which could influence the optimality of how stringent our abort conditions should be to achieve the best rates.

In summary, this thesis delivers advancements in Device Independent and semi-Device Independent randomness expansion protocols. Our contributions span enhanced protocols, tighter bounds, and the inclusion of input randomness recycling. The protocols' practical implementation depends on a delicate balance between theoretical strategies and experimental realities, an aspect that warrants further investigation.

Part III

Generalized Probability Theories

— 11 —

Generalized Probability Theories

11.1 INTRODUCTION

The Generalized Probability Theory (GPT) [93, 94] framework provides a description of physical theories based on the experimentally achievable input-output statistics of measurements, given preparations and channels. Each GPT is characterized by its state space, effect space, and set of channels. In this framework, a normalized state is represented by an element σ belonging to a convex set $\mathcal{C}$. The state space consists of tuples of the form $\omega = (\lambda, \sigma)$, where λ ranges from 0 to 1 and σ belongs to $\mathcal{C}$. The value $\lambda \in [0, 1]$ called normalization constant. An effect is defined as a linear map v on the state space satisfying $0 \leq v(\omega) \leq 1$. Physically, an effect represents an outcome of a measurement in the GPT, and its action on the state $v(\omega)$ assigns a probability to the occurrence of the event. A channel, denoted as $\mathcal{M}$, describes the dynamics of the system under specific physical processes.

Quantum theory is a particular GPT where the state space consists of density operators ρ, which are positive operators with a trace less than or equal to 1. Therefore, for quantum theory, the convex set $\mathcal{C}$ is identified as the set of non-negative operators with trace 1, and the normalization constant $\lambda = \mathrm{tr}(\rho)$. The effect space in quantum theory includes any non-negative operator E such that $\mathbb{1} - E$ is also positive. The probability of an event occurring is given by $\mathrm{tr}(\rho E)$. The set of channels $\mathcal{M}$ in quantum theory corresponds to Trace-non increasing Completely Positive Maps (TCPMs).

Checking the complete positivity requirement for a channel can be a-priori challenging, as it requires verifying the positivity of $\mathcal{M} \otimes \mathcal{I}_k$ for arbitrary natural numbers k. However, there are known results, such as Choi's theorem, that provides a tractable criterion to check whether a given linear map is a channel in quantum theory. Furthermore,

it is known that any completely positive map can always be expressed using a Kraus operator decomposition. That is, a map $\mathcal{M}$ is completely positive if and only if there exist operators $\{K_i\}_i^n$ satisfying $\sum_i^n K_i^\dagger K_i \geq 0$, such that $\mathcal{M}(\rho) = \sum_i^n K_i \rho K_i^\dagger$.

In any GPT, similar concepts of complete positivity and positivity come into play. Just like in quantum theory, where effects and channels must adhere to the notion of complete positivity, the same is true for GPTs. Computing the set of "positive" maps for most GPTs should be relatively straightforward (provided the state space can be embedded in finite dimensional Hilbert spaces), and in principle, numerical algorithms can be employed for this purpose, depending on the complexity of the state space.

However, verifying the condition for complete positivity in GPTs is a significantly more challenging task. Unlike in (finite dimensional) quantum theory, where determining whether a linear map is a channel is in principle achievable using computational methods, there are no equivalent results available for arbitrary GPTs. The main problem is that, for an arbitrary GPT, there is no unique method or principle to define composite state spaces. Furthermore, there are no existing results that allow us to express any completely positive map in an analogous Kraus operator representation. In this section of the thesis, our primary focus is to determine a tractable way to determine the set of channels for Boxworld.

Boxworld is a well-studied GPT that has attracted considerable attention [95] in the literature. In Boxworld, the state space consists of all probability distributions that do not allow super-luminal signaling. Therefore Boxworld stands as the largest theory, in terms of achievable input-output correlations, that is consistent with relativity and includes the quantum theory state space as a strict subset. The extremal points in the 2-input, 2-output state space of Boxworld are known as PR boxes [96]. These PR boxes represent non-signaling states capable of achieving a maximal CHSH score of 1, surpassing the Tsirelson bound of $\frac{1}{2} + \frac{1}{2\sqrt{2}}$ in quantum theory.

While Boxworld shares some fundamental properties with quantum mechanics, such as the no-cloning and no-broadcasting theorems, as well as the monogamy of correlations [97], it also exhibits distinct features. Notably, Boxworld allows for significantly enhanced communication power compared to quantum mechanics, as demonstrated by van Dam et al. [98] and Brassard et al. [99]. Additionally, Linden et al. [100] proved how certain post-quantum theories, including Boxworld, enable nonlocal computation tasks that are unattainable in quantum mechanics. Consequently, information-theoretic tasks have been proposed as means to distinguish Boxworld from quantum mechanics.

Even if Boxworld is the largest GPT in terms of the achievable input-output correlation,

it has very simple mathematical features. This is because Boxworld's larger state space, compared to quantum theory, significantly restricts the dynamics it can exhibit. Barrett [93] demonstrated the absence of joint measurements in Boxworld for two parties. This means that any measurement in Boxworld for two parties is essentially a local measurement, and unlike quantum theory, global measurements do not exist. Further work by Gross et al. [95] proved that reversible dynamics within Boxworld are trivial; i.e., any reversible transformation in Boxworld consists solely of local operations and permutations of systems. Notably, the ability to perform joint measurements on two systems is a fundamental aspect of phenomena like teleportation and entanglement swapping [101, 102], and hence these features of quantum theory also cannot be replicated in Boxworld. By studying Boxworld's dynamics, we can understand what sets quantum theory apart from generalized theories and gain a deeper understanding of the phenomena that arise in quantum mechanics.

In the next chapter, we carefully define Boxworld in terms of its state space. Through this definition, we develop a linear algebraic framework for Boxworld that can be applied to scenarios involving any number of non-communicating (or space-like separated) parties. Using this framework, we prove that any valid linear map that maps a Boxworld state to another state is a channel in Boxworld. Borrowing terminology from quantum theory, this means that any "positive" map in Boxworld is automatically "completely positive". This observation indicates that the set of channels in Boxworld is numerically tractable (simply check the action of the channel on the extremal Boxworld states), facilitating an efficient characterization of the dynamics of the theory.

In the remainder of the chapter, we define various aspects of the GPT framework such as states, effects and channels. Furthermore, we shall prove certain fundamental facts within the GPT framework, which apply to all GPTs.

11.2 THE GPT FRAMEWORK: A CONCISE LITERATURE OVERVIEW

Before delving into the intricacies of the GPT framework, we provide a brief overview of the literature on GPTs. The GPT framework emerged in the context of axiomatizations of quantum theory. Rather than depending on the conventional "textbook" axioms of quantum theory, which are based on abstract mathematical constructs without clear physical motivations, the goal was to formulate quantum theory with axioms from well-motivated physical principles [103–106]. To do so, the GPT framework was used to

identify the essential features a physical theory should possess and to introduce axioms to derive quantum theory from. The GPT framework appears especially appropriate for such a pursuit, as it operates under minimal requirements – it describes a physical theory solely in terms of its ability to reproduce the observable statistics of any experiment deemed physical. In other words, the GPT framework only aims to reproduce operational aspects of a theory.

The pursuit of axiomatizing quantum theory led to the conception of Boxworld. Popescu and Rohrlich, in their seminal paper [103], questioned whether the ability to violate the Bell inequality (specifically, the CHSH inequality) was unique to quantum theory. They explored whether the principles of non-locality and relativistic causality (which for the context of this work simply means that faster than speed of light communication is not possible) were sufficient to recover all experimental correlations consistent with the predictions of quantum theory. Surprisingly, they identified non-local correlations beyond the purview of quantum theory, suggesting that merely adopting non-locality and relativistic causality as axioms fails to fully reproduce quantum theory.

Barrett [93] constructed "the generalized non-signaling theory" keeping non-locality and relativistic causality two as the sole axioms. The generalized non-signaling theory [93] subsequently became known as "Boxworld". This theory encompasses all non-local correlations, including those identified by Popescu and Rohrlich, which were later called the "PR box".

Although the GPT framework was initially introduced to study quantum theory through physically motivated axioms, it has since been applied in a wide variety of contexts. For example, this framework has been extended to investigate computation [107–112] and information-theoretic tasks such as bit commitment and communication complexity [98–100, 113, 114]. An interesting question to address is the extent to which the structure of quantum theory is crucial for proving security in various Device Independent protocols. In this context, Barrett et al. [11] illustrated that the standard Device Independent Quantum Key Distribution (DIQKD) protocol remains secure against a spectrum of attacks by post-quantum eavesdroppers, who are limited solely by the non-signaling principle.

The generality of the GPT framework has led to a deep understanding of the "logical architecture" of quantum theory. A vast body of research exists that seeks to understand different features of quantum theory that are absent in classical theory. Investigating whether these features are also present in other theoretical constructs not only enhances our understanding of quantum theory but also reveals how multiple features of a theory

might be a consequence of the same logical architecture. This research offers insights into the interplay between various quantum features and the "minimal assumptions" needed for operational theories to replicate those features. Non-local features such as entanglement, Bell violation, and steering have been studied in various GPTs [115–119]. Beyond non-locality, other attributes of quantum theory, like non-contextuality, have also been studied within the GPT framework [120–122]. Additionally, the GPT framework has been used to explore theories compatible with various other physical principles and physical theories, such as causality [123], and thermodynamics [124–126]. The framework is also instrumental in addressing (apparent) logical paradoxes [127] and interpretational issues [128] in quantum theory. Furthermore, GPTs have also been used to study theories resulting from the omission of specific axioms of quantum theory [129, 130].
In summary, the quest to understand axioms for quantum theory has evolved into a vast and rich domain, allowing to study diverse operational aspects of different mathematically conceivable physical theories.

11.3 STATE SPACES

In a GPT, each system is described using a mathematical quantity known as a state. Similar to quantum states, these states provide complete information about the probability of any outcome that can occur in any given measurement performed on the system. The collection of all possible states is referred to as the state space of the GPT.
We start by defining the state space. For each system in the GPT, we associate a Hilbert space, denoted as $\mathcal{H}$. We refer dual space of $\mathcal{H}$ by the notation $\mathcal{H}^*$ (the set of all linear maps $\mathcal{H} \mapsto \mathbb{R}$). The most general state space corresponding to the system is constructed as follows:

- Identify a set $\mathcal{S}^n$ known as the set of normalized states. This set is a convex set that possesses a specific property: there exists a vector $I \in \mathcal{H}^*$ such that $\langle I, \psi \rangle = 1$ for every $\psi \in \mathcal{S}^n$. The vector I is referred to as the identity effect, and it is unique in fulfilling this property.
- Then the set $\mathcal{S}$ is simply defined as

$$\mathcal{S} = \{(\lambda, \psi) \equiv \lambda\psi| \ \ \lambda \in [0, 1], \psi \in \mathcal{S}^n\}. \tag{11.1}$$

When $\lambda < 1$, then we say that the state is un-normalized or sub-normalized. Sub-normalized states cannot represent a physical process completely, as they assign an

overall probability of less than 1 to the outcomes of any measurement. Nonetheless, they are extremely useful mathematically. Just as in quantum theory, where sub-normalized states arise naturally when considering the action of a quantum channel on a quantum state, they also emerge naturally in the GPT framework when we describe the action of a channel (defined in the next section) on a state.

In analogy with the quantum theory, for GPTs, we can define the set of pure states as the states that are not "mixed" – i.e. they cannot be a expressed as the statistical mixture of different states. The pure states are on the boundary of the state space and the state space is the convex hull of the set of pure states. We denote the set of pure states by the symbol $\partial\mathcal{S}$ (not to be confused with the boundary of the set $\mathcal{S}$)

$$\partial\mathcal{S} := \{\psi \in \mathcal{S} : \psi = \mu\psi_1 + (1-\mu)\psi_2 \implies \psi_1 = \psi_2 = \psi\}. \tag{11.2}$$

The GPT framework also allows for a complete description of the experimental statistics of multiple systems. This is done by essentially treating multiple systems as effectively a single system, with a single measurement identified as a collection of measurements on each system[1].

For a multi-system GPT capable of describing interesting phenomena, it must allow for multi-system states which produce correlated outcomes when local measurements are performed on each system. Furthermore, it is also interesting if the multi-system GPT allows for performing non-trivial "joint measurements" – i.e. a measurement that cannot be expressed as a collection of independent measurements on each system.

The way we describe a multi-system state space is by first labelling every system in our theory using a unique natural number $n \in \mathbb{N}$. We then label the single or multi-system state space using an index $i \subset \mathbb{N}$. For example, $\mathcal{S}_{\{1,2,4\}}$ represents the composite state space describing the systems 1, 2, and 4. Similarly, we use the notation $\mathcal{H}_i$ for the Hilbert space that embeds the state space $\mathcal{S}_i$. Therefore, the full description of state spaces in a GPT consists of a collection of these state spaces:

$$\mathcal{T}_{\mathcal{S}} := \{\mathcal{S}_i | i \subset \mathbb{N}\}. \tag{11.3}$$

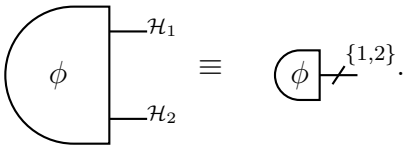
In order to ensure consistency within our theory, it is necessary for multi-system state spaces to include single-system state spaces. This means that each multi-system state

[1]There may be other types of measurements in such GPTs, however, this form of measurement must always be present.

space must contain subsets that are isomorphic to the corresponding single-system state spaces. Deferring the discussion on measurements to a later section, as is conventional in the literature, we assume the validity of local tomography for our theories. Roughly speaking, this property asserts that the global state can be fully characterized through local measurements. Under the assumption of local tomography, the constraint on single-system Hilbert spaces is given by[2] [93]

$$\mathcal{H}_{i \sqcup j} \cong \mathcal{H}_i \otimes \mathcal{H}_j.$$

Here $\otimes$ is the standard tensor product of Hilbert spaces. Unless stated otherwise, this assumption is made in this thesis. To visually represent multi-system states, we use diagrams with multiple wires, where each wire corresponds to a system's state space. For example, a two-system state $\phi \in \mathcal{S}_{\{1,2\}}$ is given by



(11.4)

The diagrammatic representation of the state space in GPTs offers a visual approach to understand different aspects of the GPT framework, which is often very useful proving important statements about the GPTs. We shall expand upon the diagrammatic version of GPTs in the last section of this chapter.

11.4 TRANSFORMATIONS

The description of a physical theory remains incomplete without a means of representing "dynamics" in systems. Within the GPT framework, such dynamics are represented using "channels".

One of the most basic behaviors a system can exhibit is remaining static, without any evolution. This lack of evolution or change is captured by a map known as the "identity channel", denoted by $\mathcal{I}$. This channel is defined as follows:

Definition 23 (Identity channel or map)**.** Let $\mathcal{S}_k$ be a state space. The identity channel $\mathcal{I}_k$ is a unique map that obeys

$$\forall \psi \in \mathcal{S}_k : \mathcal{I}_k \psi = \psi. \tag{11.5}$$

[2] Here $\sqcup$ represents the disjoint union of two sets - i.e. it is implicitly understood that sets i and j are disjoint.

The identity channel is a valid physical transformation of any GPT [94] by assumption. For majority of the GPTs, the identity map is chosen to be the map $\mathbb{1}$ on the underlying Hilbert space that embeds the state space.

Consider preparing two states, ψ_1 with probability μ and ψ_2 with probability $1-\mu$. When subjected to the physical process represented by $\mathcal{M}$, the resulting states, $\mathcal{M}\psi_1$ and $\mathcal{M}\psi_2$, should retain their probabilities as μ and $1-\mu$, respectively. In other words, a channel must respect the convex structure of the state space i.e. -

$$\forall \psi_1, \psi_2 \in \mathcal{S}, \forall \mu \in [0,1] : \mathcal{M}(\mu\psi_1 + (1-\mu)\psi_2) = \mu\mathcal{M}(\psi_1) + (1-\mu)\mathcal{M}(\psi_2). \quad (11.6)$$

On the level of the Hilbert space $\mathcal{H}$, the above condition can be fulfilled if $\mathcal{M}$ is a linear map on $\mathcal{H}$. Thus, we assume that the channels for our GPTs are linear maps acting on the Hilbert space $\mathcal{H}$.

Before formally defining a channel, we begin by defining a "transformation" or "positive map" in the context of GPTs.

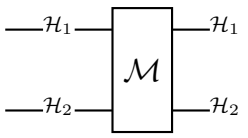
Definition 24 (Transformation or Positive map)**.** Let $\mathcal{S}_i$ be a state-space embedded in the Hilbert space $\mathcal{H}_i$, and $\mathcal{S}_j$ be a state-space embedded in the Hilbert space $\mathcal{H}_j$. Any linear map $\mathcal{M} : \mathcal{H}_i \to \mathcal{H}_j$ is called a transformation if:

$$\forall \psi \in \mathcal{S}_i : \mathcal{M}(\psi) \in \mathcal{S}_j.$$

For a transformation to be considered a channel, it must satisfy an additional requirement that it consistently acts as a transformation on composite state spaces as well. This is the analogue of the notion "complete positivity" in GPTs. We use this requirement to define channels in GPTs.

Definition 25 (Channel or completely positive map)**.** A transformation $\mathcal{M} : \mathcal{S}_i \to \mathcal{S}_j$ is a channel if for any state space $\mathcal{S}_{i \sqcup k}$, the map $\mathcal{M} \otimes \mathcal{I}_k$ is a transformation.

We will adopt the following diagrammatic representation for the transformation. For example, the transformation $\mathcal{M} : \mathcal{S}_{\{1,2\}} \mapsto \mathcal{S}_{\{1,2\}}$ is denoted by:



. (11.7)

11.5 EFFECTS

As the GPT framework aims to describe experimentally achievable statistics in any measurement, it is essential to have the ability to assign probabilities to outcomes of these measurements. These outcomes are represented by mathematical objects known as effects. When an outcome is observed in a given experiment, the probability of the outcome can be computed by action of the effect on the state in which the system is prepared. As motivated by the set of transformations, the map E must be linear in order to preserve the convexity of the state space. Thus, the effects are also taken to be linear maps. When a state resides in the space $\mathcal{H}$, the set of effects, being linear maps on the set of states, exists in the dual space $\mathcal{H}^*$.
Similar to the subtleties between channels and transformations, caution is required when defining effects. We define a witness as any linear map that assigns a probability to any state within our state space.

Definition 26 (Witness)**.** A linear map $\mathcal{H}_i \mapsto \mathbb{R}$ is a witness if

$$\forall \psi \in \mathcal{S}_i : \quad E(\psi) \in [0, 1].$$

For a witness E to correspond to an outcome of an experiment in a GPT, $E \otimes \mathcal{I}$ also needs to be a witness on all state spaces that contain the state space $\mathcal{S}$. Witnesses that can be assigned to outcomes of an experiment in a GPT are called effects.

Definition 27 (Effect)**.** A witness E is an effect, if

- for every state space $\mathcal{S}_{i \sqcup k}$, the map $E \otimes \mathcal{I}_k$ is a channel, and
- $I - E$ is an effect.

While the first condition of $E \otimes \mathcal{I}$ stems from the fact that an effect should act consistently on the composite state space, the condition that $I - E$ is also an effect is used in order to ensure an effect must form a part of a measurement in our GPT. We also assume that the set of effects, or the effect space, is a convex set. This means that any convex combination of two effects must also be an effect in the GPT.
Mathematically, it is significantly easier to check if a particular map serves as a witness than to determine if it qualifies as an effect. This is because, to verify that a linear map E is an effect, one must confirm that an infinite number of linear maps of the form $E \otimes \mathrm{id}_k$ act as witnesses. We denote the set of witnesses for the state space $\mathcal{S}_i$ by the notation $\mathcal{S}_i^*$.

This duality between states and witnesses is much stronger. Given a state space, there is a unique set of witness, however, the relation goes the other way round too – given the set of witnesses, the set

$$(\mathcal{S}^*)^* := \{\phi \in \mathcal{H} \quad | \quad \forall E \in \mathcal{S}^* : \quad E(\phi) \in [0,1]\}$$

of maps that assign probability to the witnesses is precisely the state space.

Lemma 50. *Let $\mathcal{S}$ be any state space, then*

$$(\mathcal{S}^*)^* = \mathcal{S}.$$

Proof. Let $\phi \in (\mathcal{S}^*)^*$ be such that

$$\forall E \in \mathcal{S}^* : E\phi \in [0,1],$$

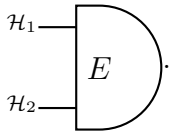
but $\phi \notin \mathcal{S}$. We prove that such a vector ϕ cannot exist. If such a vector ϕ existed, then a witness that assigns a negative probability to ϕ would also exist, which contradicts our assumption.
As $\mathcal{S}$ is a state space, it is a convex set by construction. Thus, by the separating hyperplane theorem, there exists a linear map F such that

$$F\phi \leq 0 \text{ and } F\psi \geq 0 \quad \forall \psi \in \mathcal{S}.$$

Let $\lambda = \max_{\psi \in \mathcal{S}}(F\psi)$, then clearly $\frac{1}{\lambda}F \in \mathcal{S}^*$. However $\frac{1}{\lambda}F\phi < 0$, which leads to the conclusion that $\frac{1}{\lambda}F \notin \mathcal{S}^*$, which leads to a contradiction. □

Often in GPTs, the set of states and effects are put on an equal footing. This translates to the demand that given a state space $\mathcal{S}$, the set of effects must be the set of witnesses $\mathcal{S}^*$. We call this property the **no-restriction hypothesis** [105, 131]. While this symmetry in preparations and measurements may not be crucial for an arbitrary GPT, it should be noted that this holds for quantum theory.

We also use diagrams to express the set of effects. For example, an effect E for the state space $\mathcal{S}_{\{1,2\}}$ is given by:



(11.8)

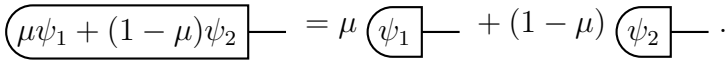
Finally, we are able to define a measurement in the GPT framework:

Definition 28. We can define a measurement in the GPT framework in terms of measurement $\{E_i\}_i$ such that $\sum_i E_i = I$.

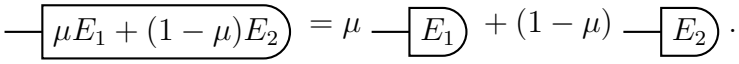
11.6 GPTS AS DIAGRAMS

The diagrammatic notation (see [94] for a full discussion and refer to papers such as [132] and [133] for examples of usage of the notation) proves to be highly convenient when dealing with GPTs and demonstrating important claims about them. Its visual nature allows us to easily identify results that might remain obscured in the conventional linear algebraic framework. A similar diagrammatic framework for finite dimensional quantum theory was introduced by Coecke et. al. [134, 135]. This graphical framework has been used for various areas, such as quantum error correction, quantum natural language processing, among other areas of quantum information theory and quantum computing [136].

Throughout the chapter, we have familiarized ourselves with the diagrammatic notation, for states, channels and effects. In light of this, let us consider expressing a state $\mu\psi_1 + (1-\mu)\psi_2$ in terms of two states ψ_1 and ψ_2, where $\mu \in [0,1]$. Employing the diagrammatic notation, this can be expressed as follows:

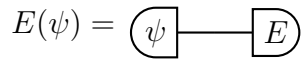


Similarly holds for effects:

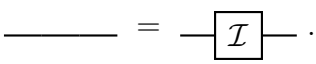


Note that above, we have dropped the label for the Hilbert space on the wires. We will omit the explicit Hilbert space label on the wires whenever the interpretation of the diagram remains unambiguous.

The primary function of GPT is to compute the probability of outcomes in any given experiment. This probability is determined by the inner product of effects and states. In the diagrammatic notation, the inner products are given by diagrams with closed legs. Let E be an effect and ψ be a state, then the probability $E(\psi)$ is given by:



Representing transformations in the diagrammatic framework is relatively straightforward. The identity channel in the GPT is denote by a wire:



To illustrate how all three components of our GPT fit together, let's consider the following example:

$$E(\mathcal{M}\psi) = \text{  } \cdot$$

As can be seen, this notation makes the linear algebraic expressions visual and thus easy to understand and perform mathematical manipulations.

— 12 —

Channels in Boxworld

This chapter is solely dedicated to exploring Boxworld, which stands as the largest theory (in terms of achievable measurement statistics) that remains consistent with the principles of relativity, as introduced in the previous chapter. This chapter begins with the definition of the most general Boxworld state space and the corresponding set of channels for Boxworld.

Following this, we delve into redefining the state space in terms of non-signaling vectors, which helps us to construct a linear-algebraic framework for Boxworld. In the final section of this chapter, we employ this framework to prove that, unlike the case with quantum theory, every transformation in Boxworld is also a channel.

12.1 NOTATIONS AND DEFINITIONS

In this section, we present the most general definition of Boxworld state space. Our aim is to establish a clear and well-defined representation of Boxworld, which will serve as the foundation for constructing a linear algebraic framework of the theory. This framework will not only facilitate our analysis but also contribute to the proof of our central result, namely the equivalence of transformations and channels in Boxworld.

The physical scenario described by Boxworld involves systems being prepared by a source and subsequently measured by spacelike-separated (or non-communicating) entities such as distinct laboratories or parties. Recall that a state in the GPT framework provides complete information about the probability of any outcome that can occur in any given measurement performed on the system. In Boxworld, a state provides the probability for every possible outcome of all the local measurements performed by the distinct parties.

A multi-system or multipartite Boxworld state can be uniquely described by conditional probability distributions of the form:

$$p_\psi(a_1, a_2, \ldots, a_n | x_1, x_2, \ldots, x_n).$$

Here, $x_i \in \{0, 1, \ldots, p_i\}$ represents the set of distinct measurements that the i^{th} party can perform, and $a_i \in \{0, 1, \ldots, k_i\}$ denotes the set of outcomes of each measurement. For convenience, we assume a fixed number of outcomes for each measurement performed by a party, although our results can be generalized to cases where this assumption does not hold. However, as the parties performing measurement are non-communicating, Boxworld can only allow for states that correspond to "non-signalling" probability distributions. For example, in the case of a two partite distribution p_ψ, the distribution is considered non-signaling (NS) if the following condition holds:

$$NS: \qquad p_\psi(a_1|x_1, x_2) = p_\psi(a_1|x_1, x_2')$$

This condition must hold for all possible values of a_1, x_1, x_2, and x_2'. In general, we introduce the set of the non-signalling distributions:

Definition 29 (Non-signalling set)**.** A probability distribution p is considered to be a non-signalling probability distribution if it satisfies the following:

$$\forall j : \sum_{a_j=0}^{k_j} p(a_1, a_2, \cdots, a_j, \cdots, a_n | x_1, x_2, \cdots, x_j, \cdots, x_n) \text{ is independent of } x_j.$$

Conditions such as the one above are called "non-signalling conditions". We denote the collection of all the probability distributions that obey the non-signalling conditions by the set $\mathcal{NS}$.

It is worth noting that the set of non-signaling probability distributions is convex, and therefore can be consistently defined to be a state space. The proof of this fact is straightforward, we shall skip the proof here for brevity.
Let us now formally define the Boxworld state space. We begin by describing a single system state space. If we label the measurements as $x \in \{0, 1, \cdots, p\}$ and assign the labels $a \in \{0, 1, \cdots, k\}$ to the outcomes of each measurement, an element in the state space of the single system (or party) Boxworld can be expressed as follows:

$$\psi = \begin{bmatrix} p_\psi(0|0) \\ \vdots \\ p_\psi(k|0) \\ - - - - \\ p_\psi(1|1) \\ \vdots \\ - - - - \\ p_\psi(0|p) \\ \vdots \\ p_\psi(k|p) \end{bmatrix} \in \mathbb{R}^{kp}. \tag{12.1}$$

Here, $p_\psi(a|x)$ represents the conditional probability of obtaining outcome a when measurement x is performed. The state in equation 12.1 is a list of kp non-negative numbers, and therefore, can be treated as an element of the vector space $\mathbb{R}^{kp}$. This means that the state space of a single system Boxworld state space is a convex set in the space $\mathbb{R}^{kp}$ for some $k, p \in \mathbb{N}$.

To facilitate a more convenient representation of Boxworld states space, we will transition to using the Dirac Notation. This transition involves using the bra-ket notation, such as $|\psi\rangle$, to represent states instead of simple Greek letters like ψ. Referring back to the state ψ as in Equation (12.1), we express it as:

$$|\psi\rangle = \sum_{a_1=0}^{k_1} \sum_{x_1=0}^{p_1} p(a_1|x_1) \left|e_{a_1|x_1}\right\rangle. \tag{12.2}$$

Here, $\{|e_{a_1|x_1}\rangle\}$ represents orthonormal basis in $\mathbb{R}^{kp}$.

This notation proves to be particularly useful as it allows us to express the multi-system state ψ in a concise manner[1]:

$$|\psi\rangle = \sum_{a}^{n} \sum_{x}^{n} p_\psi(a_1, a_2, \cdots a_n|x_1, x_2, \cdots x_n) \left|e_{a_1|x_1}\right\rangle \otimes \left|e_{a_2|x_2}\right\rangle \otimes \cdots \left|e_{a_n|x_n}\right\rangle, \tag{12.3}$$

where, $p_\psi(a_1, \cdots, a_n|x_1, \cdots, x_n)$ is the probability of obtaining the outcomes $(a_1, \cdots, a_n)$ when performing the measurement labeled by $(x_1, x_2, \cdots, x_n)$ are performed.

The underlying vector space that embeds our Boxworld state space is the tensor product space $\bigotimes_i \mathbb{R}^{k_i p_i}$, which, moving forward, will be denoted by $\mathcal{H}_N^{\mathbf{k},\mathbf{p}}$. We can now define the state space for Boxworld as follows:

[1] Here $\sum_a^n$ and $\sum_x^n$ are shorthand notations for $\sum_{a_1=0}^{k_1} \cdots \sum_{a_n=0}^{k_n}$ and $\sum_{x_1=0}^{p_1} \cdots \sum_{x_n=0}^{p_n}$ respectively.

Definition 30 (Boxworld state space)**.** Let $n \in \mathbb{N}$ denote the number of parties. Let $\mathbf{k} = (k_1, k_2, \ldots, k_n)$ denote the vector corresponding to the number of outputs for each measurement, and let $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ be the vector representing the number of distinct measurements made by each party. The state space $B(n, \mathbf{k}, \mathbf{p})$ is defined as the set of all vectors $|\psi\rangle \in \bigotimes_{i=1}^{n} \mathbb{R}^{k_i p_i}$ such that the distribution p_ψ defined by:[2]

$$p_\psi(a_1, a_2, \cdots a_n | x_1, x_2, \cdots x_n) := \left\langle e_{a_1|x_1}, e_{a_2|x_2}, \ldots, e_{a_n|x_n} \middle| \psi \right\rangle \tag{12.4}$$

is a (sub-normalized) non-signalling distribution i.e. $p_\psi = \lambda p$ for some $\lambda \in [0, 1]$ and $p \in \mathcal{NS}$.

If the parameter $\lambda = 1$, then we call the state ψ as a normalized state. For a normalized state $|\psi\rangle$, if $|\psi\rangle \in \partial B(n, \mathbf{k}, \mathbf{p})$,then we call it a "pure" state. This notation is borrowed from quantum theory, in which the pure states are the boundaries of the state space. If $\lambda < 1$, then we call the state un-normalized. In general, this parameter λ defines the norm of the state.

Definition 31 (Norm)**.** Let $|\psi\rangle \in B(n, \mathbf{k}, \mathbf{p})$, and let p_ψ be the distribution defined by (12.3). Since $p_\psi = \lambda p$ for some $p \in \mathcal{NS}$ and $\lambda \in [0, 1]$, we define the norm of $|\psi\rangle$ as $|| \, |\psi\rangle \, || = \lambda$.

The norm defined above, when acting as a map $||.|| : B(n, \mathbf{k}, \mathbf{p}) \mapsto [0, 1]$ is linear. This can be shown by observing that for any $\lambda, \epsilon \in \mathbb{R}$ and $p_1, p_2 \in \mathcal{NS}$, the distribution $\lambda p_1 + \epsilon p_2$ can be expressed as $(\lambda + \epsilon)\left(\frac{\lambda}{\lambda+\epsilon} p_1 + \frac{\epsilon}{\lambda+\epsilon} p_2\right)$. It is evident that $\frac{\lambda}{\lambda+\epsilon} p_1 + \frac{\epsilon}{\lambda+\epsilon} p_2 \in \mathcal{NS}$, which follows from the convexity of $\mathcal{NS}$. Hence, the linearity of the norm map follows from the one to one correspondence of Boxworld states and probability distributions.

Now that we have defined the state space, we can proceed to define transformations in Boxworld. Recall that transformations in GPTs are linear maps between any two state spaces in the GPT, and not every transformation is a channel. A channel is a transformation that acts consistently on composite state spaces as well. However, here we deviate from this terminology, and instead, adopt the terminology inspired from the quantum theory. Consider the following definitions:

Definition 32 (Positive maps)**.** A linear map $\mathcal{M} : \mathcal{H}_n^{\mathbf{kp}} \mapsto \mathcal{H}_{n'}^{\mathbf{k'p'}}$ is positive if

$$\forall \, |\psi\rangle \in B(n, \mathbf{k}, \mathbf{p}) : \quad \mathcal{M} \, |\psi\rangle \in B(n', \mathbf{k}', \mathbf{p}').$$

[2]The notation $|x, y\rangle$ and $|x\rangle \otimes |y\rangle$ will be used interchangeably in this context.

We introduce the notion of l positive maps

Definition 33 (l positive maps)**.** A positive linear map $\mathcal{M} : \mathcal{H}_n^{\mathbf{kp}} \mapsto \mathcal{H}_{n'}^{\mathbf{k'p'}}$ is a l positive if the map $\mathcal{M} \otimes \mathbb{1}_l : \mathcal{H}_n^{\mathbf{kp}} \otimes \mathcal{H}_l^{\tilde{\mathbf{kp}}} \mapsto \mathcal{H}_{n'}^{\mathbf{k'p'}} \otimes \mathcal{H}_l^{\tilde{\mathbf{kp}}}$ is positive. Here $\mathbb{1}_l$ is the identity defined on the space $\mathcal{H}^{\tilde{\mathbf{kp}}}$.

This inspires the definition of completely positive maps.

Definition 34 (Completely positive maps)**.** A linear map $\mathcal{M}$ is completely positive if $\mathcal{M}$ is k positive for every $k \in \mathbb{N}$.

We adopted the terminology of 'positivity' and 'complete positivity' over 'transformations' and 'channels' because it allows us to draw parallels between concepts in Boxworld and those in quantum theory.

12.2 REDEFINING BOXWORLD STATE SPACE

In this section, our objective is to find an alternative definition of the Box-world state space which is free from the direct reference to the probability distribution associated with the state. To do so, we start by introducing a special set of vectors known as the non-signaling vectors:

Definition 35 (Non-signalling vectors)**.** Let

$$|f_{x_i,x_i'}\rangle := \sum_{a_i=0}^{k_i} \left(|e_{a_i|x_i}\rangle - |e_{a_i|x_i'}\rangle \right),$$

then we define the set $v_{\mathcal{NS}}$ as the set of all the vectors of the form

$$\left| e_{a_1|x_1}, \cdots, e_{a_i|x_i}, \cdots, f_{x_j,x_j'}, \cdots, e_{a_n|x_n} \right\rangle.$$

As we will show next, the set $v_{\mathcal{NS}}$ comprises vectors that have a one-to-one correspondence with the non-signaling constraints.

Lemma 51. *Let* $|\psi\rangle \in B(n, \mathbf{k}, \mathbf{p})$*, then* $\forall\, |w\rangle \in v_{\mathcal{NS}}$*,* $\langle w|\psi\rangle = 0$ *.*

Proof. Let p_ψ be the probability distribution associated with the state ψ. Since p_ψ obeys the non-signalling constraints, we have that:

$$\sum_{a_j=0}^{k_j} \left(p_\psi(\cdots, a_j, \cdots | \cdots, x_j, \cdots x_l, \cdots) - p_\psi(\cdots, a_j, \cdots | \cdots, x_j', \cdots x_l, \cdots) \right) = 0.$$

The above can be re-written as

$$\sum_{a_j=0}^{k_j} \left(\langle e_{a_1|x_1}| \otimes \cdots \otimes \left(\langle e_{a_j|x_j}| - \langle e_{a_j|x'_j}| \right) \cdots \otimes \langle e_{a_l|x_l}| \otimes \cdots \otimes \langle e_{a_n|x_n}| \right) |\psi\rangle \quad = \quad 0.$$

This is equivalent to $\langle w|\psi\rangle = 0$ for every $|w\rangle \in v_{\mathcal{NS}}$. □

The non-signalling vectors defines a sub-space

$$\mathfrak{NS} := \text{span}\left(\{|w\rangle \,|\, |w\rangle \in v_{\mathcal{NS}}\}\right)$$

of all the vectors orthogonal to the state space. Each vector $|w\rangle \in \mathfrak{NS}$ can be used to define a different non-signalling condition, making this subspace interesting to analyse. The discussion in the section helps us to re-define Boxworld state space.

Lemma 52. *An element $|\psi\rangle \in \mathcal{H}_n^{\mathbf{kp}}$ is a state in Boxworld iff it satisfies the following:*

- ***Positive:*** *For every $a_1 \cdots a_n$ and for every $x_1 \cdots x_n$: $\langle e_{a_1|x_1}, \cdots, e_{a_n|x_n}|\psi\rangle \geq 0$.*
- ***Normalizable:*** $|| \, |\psi\rangle \, || \leq 1$.
- ***Non-signalling:*** $\forall \, |w\rangle \in \mathfrak{NS} : \langle w|\psi\rangle = 0$.

12.3 SOME EXAMPLES AND PROPERTIES OF COMPLETELY POSITIVE MAP

Now, let's consider some examples and properties of completely positive maps in Boxworld. We will specifically focus on examples that will be useful for proving our results in the next section. Let's consider the following example:

Lemma 53. *The map $\mathcal{M}_{a_i|x_i}$ defined by*

$$\mathcal{M}_{a_i|x_i} \, |\psi\rangle = \langle e_{a_i|x_i}|\psi\rangle$$

is a completely positive map.

Proof. To simplify the argument, we show that $\mathcal{M}_{a_1|x_1}$ is a completely positive map, as the reasoning is not explicitly dependent on the specific number of measurement outcomes or the total number of measurements. Let p_ψ be the probability distribution associated with the state $|\psi\rangle$ as defined in equation (12.3). We show that the distribution q defined by the components

$$q(a_2 \cdots a_n|x_2 \cdots x_n) := p_\psi(a_1, a_2 \cdots a_n|x_1, x_2 \cdots x_n) \tag{12.5}$$

is indeed a (sub-normalized) non-signaling probability distribution - i.e. q is of the form λp for some $p \in \mathcal{NS}$ and $\lambda \in [0,1]$. Note that the positivity of the components $q(a_2 \cdots a_n | x_2 \cdots x_n)$ follows trivially for the expression above. To show that q is a (sub-normalized) non-signalling probability distribution, we show that $||\mathcal{M}_{a_1|x_1} |\psi\rangle || \leq || |\psi\rangle ||$ and that q obeys all the non-signalling constraints.
We first show that $||\mathcal{M}_{a_1|x_1} |\psi\rangle || \leq || |\psi\rangle ||$ as follows:[3]

$$\begin{aligned}
\sum_{a\neq 1}^{n} q(a_2, a_3 \cdots a_n | x_2, x_3, \cdots x_n) &= \sum_{a\neq 1}^{n} p_\psi(a_1, a_2, \cdots, a_n | x_1, x_2, \cdots, x_n) \\
&\leq \sum_{a}^{n} p_\psi(a_1, a_2, \cdots, a_n | x_1, x_2, \cdots, x_n) \quad (12.6) \\
&= || |\psi\rangle ||.
\end{aligned}$$

The inequality (12.6) holds as a consequence of $p_\psi(a_1, \cdots, a_n | x_1, \cdots, x_n)$ all non-negative.
We now check if q obeys the non-signalling conditions. Observe that

$$\begin{aligned}
\sum_{a_j=0}^{k_j} q(\cdots, a_j, \cdots | x_2, \cdots, x_j, \cdots) &= \sum_{a_j=1}^{k_j} p_\psi(a_1, \cdots, a_j, \cdots | x_1 \cdots, x_j, \cdots) \\
&= \sum_{a_j=0}^{k_j} p_\psi(a_1, \cdots, a_j, \cdots | x_1 \cdots, x'_j, \cdots,) \\
&= \sum_{a_j=0}^{k_j} q(\cdots, a_j, \cdots | x_2, \cdots, x'_j, \cdots).
\end{aligned}$$

Importantly, the validity of our argument is independent of the number of parties and the number of measurements each party makes, meaning that $\mathcal{M}$ is completely positive. □

By utilizing distinct positive maps in Boxworld, we can construct other completely positive maps. One straightforward way to achieve this is by composing positive maps. If we have two completely positive maps $\mathcal{M}$ and $\mathcal{N}$, their composition $\mathcal{M} \circ \mathcal{N}$ is also a completely positive map. Consequently, the composition of multiple maps of the form $\mathcal{M}_{a_i|x_i}$ - i.e. maps of the form

$$\mathcal{M}_{a_1|x_1} \circ \mathcal{M}_{a_2|x_2} \circ \cdots \circ \mathcal{M}_{a_j|x_j}$$

is a completely positive map.

[3] The notation $\sum_{a\neq 1}^{n}$ is a short hand notation for $\sum_{a_2=0}^{k_2} \cdots \sum_{a_3=0}^{k_3} \cdots \sum_{a_n}^{k_n}$.

In some cases, summing two completely positive maps can result in the construction of a new completely positive map. To see why this is the case, let $|\psi\rangle$ be a Boxworld state. If $\mathcal{M}$ and $\mathcal{N}$ are both completely positive maps, then the vector $(\mathcal{M}+\mathcal{N})|\psi\rangle$ is orthogonal to the set $\mathfrak{MG}$, which can be easily seen by taking the appropriate inner products. Therefore, to establish whether $(\mathcal{M}+\mathcal{N})|\psi\rangle$ is a valid state, we only need to verify if $||(\mathcal{M}+\mathcal{N})|\psi\rangle|| \leq 1$ holds for every state $|\psi\rangle$.
One such example is the map $\mathcal{M}_{x_1}$ defined by

$$\mathcal{M}_{x_i}|\psi\rangle = \langle e_{x_i}|\psi\rangle.$$

where $|e_{x_i}\rangle := \sum_{a_i=0}^{k_i} \left|e_{a_i|x_i}\right\rangle$. From the discussion above, to show that this map is completely positive, it suffices to show that $\mathcal{M}_{x_i}$ is norm non-increasing. In fact, we show that this map preserves the norm of the state:

$$\begin{aligned}
||\mathcal{M}_{x_1}|\psi\rangle|| &= \sum_{a_1=0}^{k_1} ||\mathcal{M}_{a_1|x_1}|\psi\rangle|| \\
&= \sum_{a_1=0}^{k_1}\sum_{a\neq 1}^{n} p_\psi(a_1,a_2,\cdots,a_n|x_1,x_2,\cdots,x_n) \\
&= ||\,|\psi\rangle\,||.
\end{aligned}$$

Here, we have used the linearity of the norm map. Therefore, the map $\mathcal{M}_{x_1}$ is completely positive.
Again, it is worth mentioning explicitly that the composition of such maps such as

$$\mathcal{M}_{x_1}\circ\mathcal{M}_{x_2}\circ\cdots\circ\mathcal{M}_{x_n}$$

yields a completely positive map. Moreover, this composition of maps individually preserves the norm of the state. Thus, we can define the norm map in terms of such composition - i.e. for every $\psi \in B(n,\mathbf{k},\mathbf{p})$:

$$\|\,|\psi\rangle\| = \|\mathcal{M}_{x_1}\circ\mathcal{M}_{x_2}\circ\cdots\circ\mathcal{M}_{x_n}|\psi\rangle\|.$$

Observe that $\mathcal{M}_{x_i}$ does not increase the norm of the state. We will now prove a more general claim: any transformation (positive map) in a GPT cannot increase the norm of the state. Here, we prove this claim in context of Boxworld, the arguments employed are applicable to any GPT.

Lemma 54. *Let $\mathcal{M} : B(n,\mathbf{k},\mathbf{p}) \mapsto B(n',\mathbf{k}',\mathbf{p}')$ be a positive map. Then, for all $|\psi\rangle \in B(n,\mathbf{k},\mathbf{p})$, we have $||\,|\psi\rangle\,|| \geq ||\mathcal{M}|\psi\rangle||$.*

Proof. Assume the existence of $|\psi\rangle \in B(n, \mathbf{k}, \mathbf{p})$ such that $|| \, |\psi\rangle \, || < ||\mathcal{M} \, |\psi\rangle \, || \leq 1$. Choose $\lambda \in \left(\frac{1}{||\mathcal{M}|\psi\rangle||}, \frac{1}{|||\psi\rangle||}\right)$. Clearly, $||\lambda \, |\psi\rangle \, || < 1$, indicating that $\lambda \, |\psi\rangle \in B(n, \mathbf{k}, \mathbf{p})$. However, we observe that $||\mathcal{M}(\lambda \, |\psi\rangle)|| = ||\lambda \mathcal{M} \, |\psi\rangle \, || \geq 1$. Therefore, $\mathcal{M}(\lambda \, |\psi\rangle) \notin B(n', \mathbf{k}', \mathbf{p}')$. As a result, $\mathcal{M}$ is not a positive map, leading to a contradiction. □

Note that replacing $B(n, \mathbf{k}, \mathbf{p})$ by $\mathcal{S}_n$ and $B(n', \mathbf{k}', \mathbf{p}')$ by $\mathcal{S}_m$ proves the claim for general GPTs.

12.4 POSITIVE MAPS AND COMPLETELY POSITIVE MAPS IN BOXWORLD

We are now prepared to prove the central claim of this chapter, which asserts that any positive map in Boxworld is also completely positive. Recall that to establish a vector as a state in Boxworld, it is sufficient to prove that the vector possesses three properties: it has non-negative entries in the canonical bases, it has a norm less than or equal to 1, and it is orthogonal to the set $\mathfrak{MG}$. We will break our proof into three steps, wherein we demonstrate that $\mathcal{M} \otimes \mathbb{1}_l \, |\psi\rangle$, for any $l \in \mathbb{N}$, satisfies all three conditions when $\mathcal{M}$ is a positive map and $|\psi\rangle$ is a Boxworld state (in the domain of $\mathcal{M} \otimes \mathbb{1}_l \, |\psi\rangle$). Let us begin by proving that $\mathcal{M} \otimes \mathbb{1}_l \, |\psi\rangle$ exhibits non-negative components in all canonical bases for every Boxworld state $|\psi\rangle$ in the domain of $\mathcal{M} \otimes \mathbb{1}_l \, |\psi\rangle$.

Lemma 55. *Let $\mathcal{M} : \mathcal{H}_n^{\mathbf{kp}} \mapsto \mathcal{H}_{n'}^{\mathbf{k'p'}}$ be a positive map, and let $|\psi\rangle \in \mathcal{H}_n^{\mathbf{kp}} \otimes \mathcal{H}_{n'}^{\tilde{\mathbf{kp}}}$ be any Boxworld, state then*

$$\langle e_{a_1|x_1}, \cdots, e_{a_{n'}|x_{n'}}, \cdots, e_{a_{n'+l}|x_{n'+l}} | \mathcal{M} \otimes \mathbb{1}_l | \psi \rangle \geq 0 \tag{12.7}$$

holds for every possible of measurement $(x_0, x_1, \cdots x_{n'+l})$ and output $(a_1, a_2, \cdots a_{n'+l})$

.

Proof. Let $|\psi\rangle \in \mathcal{H}_n^{\mathbf{kp}} \otimes \mathcal{H}_l^{\mathbf{k'p'}}$ be a Boxworld state. Notice that the inner product

$$\langle e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}| \otimes \langle e_{a_{n'+1}|x_{n'+1}}, \cdots, e_{a_{n'+l}|x_{n'+l}} | \mathcal{M} \otimes \mathbb{1}_l | \psi \rangle$$

can be rewritten as

$$\langle e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}} | \mathcal{M} \left(\langle e_{a_{n'+1}|x_{n'+1}}, \cdots e_{a_{n'+l}|x_{n'+l}} | \psi \rangle \right).$$

By Lemma 53, the vector $\left|\tilde{\psi}\right\rangle$ defined by $\left(\langle e_{a_{n'+1}|x_{n'+1}}, \cdots e_{a_{n'+l}|x_{n'+l}}|\psi\rangle\right)$ is indeed an n-partite state, i.e., $\left|\tilde{\psi}\right\rangle = \left(\langle e_{a_{n'+1}|x_{n'+1}}, \cdots e_{a_{n'+l}|x_{n'+l}}|\psi\rangle\right) \in B(n, \mathbf{k}, \mathbf{p})$. This is because

$$\langle e_{a_{n'+1}|x_{n'+1}}, \cdots, e_{a_{n'+l}|x_{n'+l}}|\psi\rangle = \mathcal{M}_{a_{n'+1}|x_{n'+1}} \circ \mathcal{M}_{a_{n'+2}|x_{n'+2}} \circ \cdots \circ \mathcal{M}_{a_{n'+l}|x_{n'+l}} |\psi\rangle$$

is a vector that is obtained by repeated application of positive maps and thus results in a state in the set $B(n, \mathbf{k}, \mathbf{p})$.
Now, by definition, $\langle e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}|\mathcal{M}|\tilde{\psi}\rangle \in B(n', \mathbf{k}', \mathbf{p}')$ since $\mathcal{M}$ is a positive map. Thus,

$$\langle e_{a_1|x_1}, \cdots, e_{a_{n'}|x_{n'}}, \cdots, e_{a_{n'+l}|x_{n'+l}}|\mathcal{M} \otimes \mathbb{1}_l|\psi\rangle \geq 0.$$

□

In the second part of our proof, we face the most challenging aspect of the overall argument. We show that $\mathcal{M} \otimes \mathbb{1}_l$ cannot be a signalling map. This means that the resulting vector, obtained by applying $\mathcal{M} \otimes \mathbb{1}_l$ to a state in Boxworld, is orthogonal to the subspace $\mathfrak{NS}$.

Lemma 56. *If $\mathcal{M} : \mathcal{H}_n^{\mathbf{kp}} \mapsto \mathcal{H}_{n'}^{\mathbf{k'p'}}$ is positive, then $\forall\, |w\rangle \in v_{\mathcal{NS}}$,*

$$\langle w|\mathcal{M} \otimes \mathbb{1}_l|\psi\rangle = 0$$

holds for all Boxworld states $|\psi\rangle \in \mathcal{H}_n^{\mathbf{kp}} \otimes \mathcal{H}_l^{\tilde{\mathbf{kp}}}$.

Proof. As we know that $\mathcal{M}$ is a positive map, the following holds for every $|\phi\rangle \in B(n, \mathbf{k}, \mathbf{p})$ and for every $|w\rangle \in \mathfrak{NS}$,

$$\langle w|\, \mathcal{M}\, |\phi\rangle = 0.$$

The only way the equation above holds for every $|\phi\rangle \in B(n, \mathbf{k}, \mathbf{p})$ is when $\mathcal{M}^\dagger$ leaves the space $\mathfrak{NS}$ invariant - i.e. $\forall\, |w\rangle \in \mathfrak{NS} : \mathcal{M}^\dagger\, |w\rangle \in \mathfrak{NS}$.
We need to check if for every $|w\rangle \in \mathfrak{NS}$, $\langle w|\, (\mathcal{M} \otimes \mathbb{1}_l)\, |\psi\rangle = 0$. As the map $\mathcal{M}$ acts on the first n tensor factors of the state $|\psi\rangle$, the vector $\mathcal{M}^\dagger$ on $|w\rangle$ acts on the first n' tensor factors in the dual space. This allows us to consider 2 distinct types of vectors $|w\rangle$ (up to relabeling of parties)

$$\textbf{Case 1} : |v_1\rangle \;=\; \left|f_{x_1,x_1'}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}\right\rangle \otimes \left|e_{a_{n'+1}|x_{n'+1}}, \cdots, e_{a_{n'+l}|x_{n'+l}}\right\rangle.$$

$$\textbf{Case 2} : |v_2\rangle \;=\; \left|e_{a_1|x_1}, e_{a_2|x_2}, .., e_{a_{n'}|x_{n'}}\right\rangle \otimes \left|f_{x_{n'+1},x'_{n'+1}}, e_{a_{n'+2}|x_{n'+2}}, \cdots, e_{a_{n'+l}|x_{n'+l}}\right\rangle.$$

We now deal with both the cases individually.

Case 1: We can write

$$\mathcal{M}^\dagger \otimes \mathbb{1}_l \left|v_1\right\rangle \;=\; (\mathcal{M}^\dagger \left|f_{x_1,x_1'}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}\right\rangle) \otimes \left|e_{a_{n'+1}|x_{n'+1}}, \cdots, e_{a_{n'+l}|x_{n'+l}}\right\rangle.$$

Here, $\mathcal{M}^\dagger \left|f_{x_1,x_1'}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}\right\rangle \in \mathfrak{NS}$ as discussed above. Thus, $(\mathcal{M} \otimes \mathbb{1}_l)\left|v_1\right\rangle$ must be expressable as a linear combination of vectors of the form

$$|w\rangle \otimes |e_{a_{n'+1}|x_{n'+1}}, \cdots, e_{a_{n'+l}|x_{n'+l}}\rangle,$$

where $|w\rangle \in v_{\mathcal{NS}}$. This shows that $\langle v_1|\mathcal{M} \otimes \mathbb{1}_l|\psi\rangle = 0$.

Alternatively, we can prove $\langle v_1|\,\mathcal{M} \otimes \mathbb{1}_l\,|\psi\rangle = 0$ by observing that $\langle v_1|\,\mathcal{M} \otimes \mathbb{1}_l\,|\psi\rangle$ can be written as:

$$\langle v_1|\mathcal{M}|\psi\rangle = \langle (f_{x_1,x_1'}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}|\mathcal{M}\left(\left\langle e_{a_{n'+1}|x_{n'+1}}, \cdots, e_{a_{n'+l}|x_{n'+l}}\right|\psi\right\rangle\right)$$

However, notice from Lemma 53, we can infer that

$$\left|\tilde{\psi}\right\rangle := \left\langle e_{a_{n'+1}|x_{n'+1}}, e_{a_{n'+2}|x_{n'+2}}, \cdots, e_{a_{n'+l}|x_{n'+l}}\right|\psi\right\rangle \in B(n, \mathbf{k}, \mathbf{p}),$$

as it is obtained as a composition of positive maps acting on a Boxworld state. Thus,

$$\langle v_1|\,\mathcal{M} \otimes \mathbb{1}_l\,|\psi\rangle = \langle f_{x_1,x_1'}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}|\tilde{\psi}\rangle = 0,$$

since $|f_{x_1,x_1'}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}\rangle \in \mathfrak{NS}$.

Case 2: Notice that, vectors of the form $\left|e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}\right\rangle$ span the entire space $\mathcal{H}_{n'}^{\mathbf{k}'\mathbf{p}'}$. Thus, we can express $\mathcal{M}^\dagger \left|e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_{n'}|x_{n'}}\right\rangle$ using a linear combination:

$$\sum_{a}^{n'}\sum_{x}^{n'} \alpha_{a_1,a_2,\cdots,a_{n'};x_1,x_2,\cdots x_{n'}} \left|e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_n|x_{n'}}\right\rangle,$$

where $\alpha_{a_1,a_2,\cdots a_{n'};x_1,x_2,\cdots x_{n'}} \in \mathbb{R}$ are some real coefficients. Thus,

$$\mathcal{M}^\dagger \otimes \mathbb{1}_l\,|v_2\rangle = \sum_{a}^{n'}\sum_{x}^{n'} \alpha_{a_1,a_2,\cdots,a_{n'};x_1,x_2,\cdots x_{n'}} \left|e_{a_1|x_1}, e_{a_2|x_2}, \cdots, e_{a_n|x_{n'}}\right\rangle \otimes |w\rangle,$$

where $|w\rangle \in v_{\mathcal{NS}}$. Thus, $\mathcal{M}^\dagger \otimes \mathbb{1}_l\,|v_2\rangle \in \mathfrak{NS}$. This implies that $\langle v_2|\,\mathcal{M} \otimes \mathbb{1}_l\,|\psi\rangle = 0$. Combining the results for Case 1 and Case 2, we can prove the claim. □

We now end the proof by showing that $\mathcal{M} \otimes \mathbb{1}_k$ cannot increase the norm of the states.

Lemma 57. *If $\mathcal{M} : \mathcal{H}_n^{\mathbf{kp}} \mapsto \mathcal{H}_{n'}^{\mathbf{k'p'}}$ is positive, then $\mathcal{M} \otimes \mathbb{1}_k$ is norm non-increasing.*

Proof. Let $|\psi\rangle \in \mathcal{H}_n^{\mathbf{kp}} \otimes \mathcal{H}_l^{\tilde{\mathbf{kp}}}$ be a state. From our discussion in the previous section, we can express $||\mathcal{M} \otimes \mathbb{1}_l |\psi\rangle||$ as

$$\langle e_{x_1}, e_{x_2}, \cdots e_{x_{n'}} | \mathcal{M}(\langle e_{x_{n'+1}}, \cdots, e_{x_{n'+l}} | \psi \rangle) \equiv \langle e_{x_1}, e_{x_2}, \cdots e_{x_{n'}} | \mathcal{M} | \hat{\psi} \rangle)$$

for some Boxworld state $|\hat{\psi}\rangle := \langle e_{x_{n'+1}}, \cdots, e_{x_{n'+l}} | \psi \rangle$. Note that $|\hat{\psi}\rangle$ is a Boxworld state as

$$\langle e_{x_{n'+1}}, \cdots, e_{x_{n'+l}} | \psi \rangle = \mathcal{M}_{x_{n'+1}} \circ \cdots \circ \mathcal{M}_{x_{n'+l}} |\psi\rangle = |\hat{\psi}\rangle.$$

Using above and discussion from the previous section, we can also infer that $|||\hat{\psi}\rangle|| = ||\,|\psi\rangle\,||$. From above, we know that,

$$\begin{aligned} ||\mathcal{M} \otimes \mathbb{1}_l |\psi\rangle|| &= \langle e_{x_1}, \cdots, e_{x_{n'}} | \mathcal{M} | \hat{\psi} \rangle \\ &= ||\mathcal{M} \left|\hat{\psi}\right\rangle|| \\ &\leq ||\left|\hat{\psi}\right\rangle|| \\ &= ||\,|\psi\rangle\,||. \end{aligned}$$

The inequality above follows from the Lemma 54. □

12.5 CONCLUSION

In this chapter, we began by introducing a general definition of Boxworld for arbitrary many parties. We briefly discussed what non-signalling probability distributions are and defined an appropriate state-space for both single-party and multi-partite Boxworld state spaces. Subsequently, we re-defined the Boxworld state space using a linear algebraic framework that does not directly reference the probability distributions associated with a state. Instead, it is described using three different linear algebraic constraints that are equivalent to verifying if the underlying probability distribution is a non-signalling probability distribution. We also explored some straightforward examples of completely composite maps in Boxworld and discussed some easily provable properties of completely positive maps. Using these examples and the three alternative criteria, we proved that any positive map in Boxworld is also completely positive.

It remains an open problem to identify all theories in which such a straightforward result is true. Additionally, it's interesting to determine if all the GPTs that abide by such a result also indeed lack non-trivial channels, as is the case in Boxworld. We discuss some of these problems, as well as some (partial) progress in this direction, in the next and final chapter of the thesis.

— 13 —

Discussion and conclusion

In this section of the thesis, we delve into the GPT framework, starting with the definitions of states, effects, and channels. Given the analogous concept of complete positivity in a GPT, we highlight the mathematical challenges associated with defining a channel and an effect when a GPT is determined solely by its state space. We then turn our attention to Boxworld, the theory that is capable of producing any multi-partite non-signaling probability distributions. Importantly, we make no assumptions about the number of parties involved, the number of measurements each party can conduct, or the number of outputs for each measurement.

Our definition of Boxworld results in a fully linear algebraic framework for the theory. Contrasting with quantum theory, a distinction arises: in Boxworld, any transformation (positive map) is a channel (completely positive map). This insight provides a tractable criterion, in principle, for determining whether a particular linear map qualifies as a channel within Boxworld.

Boxworld's limited dynamics have lead to variations like Houseworld [116, 137] and Noisy Boxworld [131] to capture specific quantum theory features absent in classical theory. Houseworld emerged to address the absence of non-locality swapping in Boxworld, an analog of entanglement swapping in quantum theory. In Houseworld, specific effects demonstrate non-locality swapping, illustrating that this phenomenon isn't exclusive to quantum theory. The two-party state space in Houseworld is defined by excluding certain extremal states (PR boxes) from Boxworld's state space.

Noisy Boxworld replicates various quantum theory aspects, especially Tsirelson's bound for maximally entangled states. The state space (or the effect space) of this theory is constructed by taking a convex hull of "noisy" versions of extremal states (or effects) of Boxworld.

Both Houseworld and certain Noisy Boxworld instances lack specific relabeling symmetries. Preliminary research (in collaboration with Mr. Kuntal Sengupta) suggests tractable criteria for defining effects in these theories when using the min-tensor product. Notably, the no-restriction hypothesis appears inconsistent with the max tensor product in these contexts.

This indicates a need to delve deeper into discerning sets of channels and effects, as such determinations can be intricate. Addressing such questions is important for understanding non-trivial GPTs, especially those being designed to study interesting properties of multi-partite state spaces.

Ultimately, one of the important pursuit of understanding GPTs is to axiomatize quantum theory from an information-theoretical perspective. A good method for defining the effects and channels for GPTs allowing for composite state spaces will play a crucial role if our goal is to understand all non-classical features of quantum theory using the GPT framework.

— A —

Appendix for Device Independent Protocols

A.1 UPPER BOUNDING THE DERIVATIVES - H(A|XYE)

In this section, we derive bounds on the functions $\cos(u(\mathbf{x}) \pm v)\sin(2\theta)$. Since $\sin(2\theta)$ is always positive and increasing in θ in our domain, we get the following result for any $(\eta, \theta, v) \in \mathcal{C}_{i,j,k}$. We can lower bound $\cos(u(\mathbf{x}) \pm v)\sin(2\theta)$ as

$$\begin{cases} \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left(\cos(u(\mathbf{x}) \pm v)\right) \sin(2\theta_{i+1}) & \text{if} \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left(\cos(u(\mathbf{x}_{i,j,k}) \pm v)\right) > 0 \\ \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left(\cos(u(\mathbf{x}) \pm v)\right) \sin(2\theta_i) & \text{if} \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left(\cos(u(\mathbf{x}_{i,j,k}) \pm v)\right) < 0. \end{cases} \tag{A.1}$$

Let $\mathbf{x} = (\eta, \theta, v) \in \mathcal{C}_{i,j,k}$ and, for brevity, write $g_{\pm,y}(\mathbf{x}) = u_\pm(\mathbf{x}) + (-1)^y v$ with $y \in \{0, 1\}$. Then, by Taylor's theorem (cf. Theorem 3), there exists $\mathbf{x}' \in \mathcal{C}_{i,j,k}$ such that

$$\begin{aligned} \cos(g_{\pm,y}(\mathbf{x})) \;=\; & \cos(g_{\pm,y}(\mathbf{x}_{i,j,k})) + \partial_\eta \cos(g_{\pm,y}(\mathbf{x}))\Big|_{\mathbf{x}'}(\eta - \eta_i) + \\ & \partial_\theta \cos(g_{\pm,y}(\mathbf{x}))\Big|_{\mathbf{x}'}(\theta - \theta_j^{(i)}) + \partial_v \cos(g_{\pm,y}(\mathbf{x}))\Big|_{\mathbf{x}'}(v - v_k^{(i,j)}). \end{aligned} \tag{A.2}$$

We upper bound this by upper bounding each of the partial derivatives on $\mathcal{C}_{i,j,k}$:

$$\begin{aligned} \partial_\eta \cos(g_{\pm,y}(\mathbf{x})) &= -\sin(g_{\pm,y}(\mathbf{x}))\partial_\eta u_\pm \\ \partial_\theta \cos(g_{\pm,y}(\mathbf{x})) &= -\sin(g_{\pm,y}(\mathbf{x}))\partial_\theta u_\pm \\ \partial_v \cos(g_{\pm,y}(\mathbf{x})) &= -\sin(g_{\pm,y}(\mathbf{x}))(\partial_v u_\pm + (-1)^y). \end{aligned}$$

We have bounded the derivatives of u on $\mathcal{C}_{i,j,k}$ in (5.12).

We now consider the different cases. Firstly, suppose $\max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} [-\sin(g_{\pm,y}(\mathbf{x}))] \geq 0$.

Consider the terms in (A.2). Using the bounds in (5.12), we have

$$\begin{aligned}
\partial_\eta \cos(g_{+,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\eta-\eta_i) &\leq 2 \max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{+,y}(\mathbf{x}))\right](\eta_{i+1}-\eta_i) \\
\partial_\theta \cos(g_{+,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\theta-\theta_j^{(i)}) &\leq 0 \\
\partial_v \cos(g_{+,y}(\mathbf{x}))\big|_{\mathbf{x}'}(v-v_k^{(i,j)}) &\leq \begin{cases} 2\max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{+,0}(\mathbf{x}))\right](v_{k+1}^{(i,j)}-v_k^{(i,j)}) & y=0 \\ 0 & y=1 \end{cases}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\partial_\eta \cos(g_{-,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\eta-\eta_i) &\leq 0 \\
\partial_\theta \cos(g_{-,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\theta-\theta_j^{(i)}) &\leq 0 \\
\partial_v \cos(g_{-,y}(\mathbf{x}))\big|_{\mathbf{x}'}(v-v_k^{(i,j)}) &\leq \begin{cases} 2\max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{-,0}(\mathbf{x}))\right](v_{k+1}^{(i,j)}-v_k^{(i,j)}) & y=0 \\ 0 & y=1 \end{cases}.
\end{aligned}$$

Combining all of these, and bounding the $-\sin(g_{\pm,y}(\mathbf{x}))$ terms by 1, we find

$$\begin{aligned}
\cos(g_{+,0}(\mathbf{x})) &\leq \cos(g_{+,0}(\mathbf{x}_{i,j,k})) + 2(\eta_{i+1}-\eta_i) + 2(v_{k+1}^{(i,j)}-v_k^{(i,j)}) \\
\cos(g_{+,1}(\mathbf{x})) &\leq \cos(g_{+,1}(\mathbf{x}_{i,j,k})) + 2(\eta_{i+1}-\eta_i) \\
\cos(g_{-,0}(\mathbf{x})) &\leq \cos(g_{-,0}(\mathbf{x}_{i,j,k})) + 2(v_{k+1}^{(i,j)}-v_k^{(i,j)}) \\
\cos(g_{-,1}(\mathbf{x})) &\leq \cos(g_{-,1}(\mathbf{x}_{i,j,k}))
\end{aligned} \tag{A.3}$$

Secondly, in the case $\max_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{\pm,y}(\mathbf{x}))\right] \leq 0$, we have

$$\begin{aligned}
\partial_\eta \cos(g_{+,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\eta-\eta_i) &\leq 0 \\
\partial_\theta \cos(g_{+,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\theta-\theta_j^{(i)}) &\leq -2 \min_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{\pm,y}(\mathbf{x}))\right](\theta_{j+1}^{(i)}-\theta_j^{(i)}) \\
\partial_v \cos(g_{+,y}(\mathbf{x}))\big|_{\mathbf{x}'}(v-v_k^{(i,j)}) &\leq \begin{cases} 0 & y=0 \\ -\min_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{+,0}(\mathbf{x}))\right](v_{k+1}^{(i,j)}-v_k^{(i,j)}) & y=1 \end{cases}.
\end{aligned}$$

and

$$\begin{aligned}
\partial_\eta \cos(g_{-,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\eta-\eta_i) &\leq -2 \min_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{\pm,y}(\mathbf{x}))\right](\eta_{i+1}-\eta_i) \\
\partial_\theta \cos(g_{-,y}(\mathbf{x}))\big|_{\mathbf{x}'}(\theta-\theta_j^{(i)}) &\leq -2 \min_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{\pm,y}(\mathbf{x}))\right](\theta_{j+1}^{(i)}-\theta_j^{(i)}) \\
\partial_v \cos(g_{-,y}(\mathbf{x}))\big|_{\mathbf{x}'}(v-v_k^{(i,j)}) &\leq \begin{cases} 0 & y=0 \\ -\min_{\mathbf{x}\in\mathcal{C}_{i,j,k}} \left[-\sin(g_{+,0}(\mathbf{x}))\right](v_{k+1}^{(i,j)}-v_k^{(i,j)}) & y=1 \end{cases}.
\end{aligned}$$

Combining all of these, and bounding the $-\sin(g_{\pm,y}(\mathbf{x}))$ terms by -1, we find

$$\begin{aligned}
\cos(g_{+,0}(\mathbf{x})) &\leq \cos(g_{+,0}(\mathbf{x}_{i,j,k})) + 2(\theta^{(i)}_{j+1} - \theta^{(i)}_{j}) \\
\cos(g_{+,1}(\mathbf{x})) &\leq \cos(g_{+,1}(\mathbf{x}_{i,j,k})) + 2(\theta^{(i)}_{j+1} - \theta^{(i)}_{j}) + (v^{(i,j)}_{k+1} - v^{(i,j)}_{k}) \\
\cos(g_{-,0}(\mathbf{x})) &\leq \cos(g_{-,0}(\mathbf{x}_{i,j,k})) + 2(\eta_{i+1} - \eta_{i}) + 2(\theta^{(i)}_{j+1} - \theta^{(i)}_{j}) \\
\cos(g_{-,1}(\mathbf{x})) &\leq \cos(g_{-,1}(\mathbf{x}_{i,j,k})) + 2(\eta_{i+1} - \eta_{i}) + 2(\theta^{(i)}_{j+1} - \theta^{(i)}_{j}) + (v^{(i,j)}_{k+1} - v^{(i,j)}_{k}).
\end{aligned} \tag{A.4}$$

A.2 UPPER BOUNDING THE DERIVATIVES - H(AB|00E)

For brevity in this section we often use $\bar{\theta} = \pi/4 + \theta$. We upper-bound the derivatives for the functions $\hat{\alpha}_0, \tilde{\epsilon}, \hat{R}$. We first upper bound the derivatives for α as

$$\left|\partial_\lambda \hat{\alpha}_0\right| = \left|\frac{2\cot\left(\bar{\theta}\right)\csc^2(\lambda)}{\cot^2\left(\bar{\theta}\right)\cot^2(\lambda)+1}\right| \tag{A.5}$$

$$\left|\partial_v \hat{\alpha}_0\right| = \left|\frac{\cot\left(\bar{\theta}\right)\csc^2(v)}{\cot^2\left(\bar{\theta}\right)\cot^2(v)+1}\right| \tag{A.6}$$

$$\left|\partial_{\bar{\theta}} \hat{\alpha}_0\right| = \left|\frac{2\csc^2\left(\bar{\theta}\right)\cot(\lambda)}{\cot^2\left(\bar{\theta}\right)\cot^2(\lambda)+1} - \frac{\csc^2\left(\bar{\theta}\right)\cot(v)}{\left(\cot^2\left(\bar{\theta}\right)\cot^2(v)+1\right)}\right|. \tag{A.7}$$

Observe that for $x \in \mathbb{R}$

$$\frac{a\csc^2(x)}{a^2\cot^2(x)+1} \leq \max\{a, \frac{1}{a}\}. \tag{A.8}$$

Noting that $\cot\left(\bar{\theta}\right) \leq 1$ for $\theta \in [0, \frac{\pi}{4}]$. This gives us

$$|\partial_\lambda \hat{\alpha}_0| \leq 2\tan\left(\bar{\theta}\right) =: \alpha_\lambda \tag{A.9}$$

$$|\partial_v \hat{\alpha}_0| \leq \tan\left(\bar{\theta}\right) =: \alpha_v. \tag{A.10}$$

The identity $\frac{x}{1+a^2x^2} \leq \frac{1}{2|a|}$ can be used to get the following upper bound

$$\begin{aligned}
\left|\partial_{\bar{\theta}} \hat{\alpha}_0\right| &\leq \left|\frac{2\csc^2\left(\bar{\theta}\right)\cot(\lambda)}{\cot^2\left(\bar{\theta}\right)\cot^2(\lambda)+1}\right| + \left|\frac{\csc^2\left(\bar{\theta}\right)\cot(v)}{\left(\cot^2\left(\bar{\theta}\right)\cot^2(v)+1\right)}\right| \\
&\leq 2\frac{\csc^2(\bar{\theta})\tan\left(\bar{\theta}\right)}{2} + \frac{\csc^2(\bar{\theta})\tan\left(\bar{\theta}\right)}{2} \\
&= \frac{3}{2\sin\left(2\bar{\theta}\right)} := \alpha_{\bar{\theta}}.
\end{aligned}$$

Define $z(\lambda, v, \theta)$ to be the denominator in (5.31), i.e.,

$$z(\lambda,v,\theta) := \cos(v-\lambda)\left[\sin(\hat{\alpha}_0)\sin(v)\cos\left(\frac{\pi}{4}+\theta\right)+\cos(\hat{\alpha}_0)\cos(v)\sin\left(\frac{\pi}{4}+\theta\right)\right] - \frac{\sin(v-\lambda)}{\sqrt{2}}\sqrt{1-\cos(2v)\sin(2\theta)}. \quad \text{(A.11)}$$

We now compute the derivatives of z. For the derivative with respect to λ, we write $\partial_\lambda z = (b_1 + b_3) + b_2\partial_\lambda\hat{\alpha}_0$, where

$$\begin{aligned} b_1 &= \sin(v-\lambda)\left(\sin(\hat{\alpha}_0)\sin(v)\cos\left(\bar{\theta}\right)+\cos(\hat{\alpha}_0)\cos(v)\sin\left(\bar{\theta}\right)\right) \\ b_2 &= \cos(v-\lambda)\left(\cos(\hat{\alpha}_0)\sin(v)\cos\left(\bar{\theta}\right)-\sin(\hat{\alpha}_0)\cos(v)\sin\left(\bar{\theta}\right)\right) \\ b_3 &= \frac{\cos(v-\lambda)\sqrt{1-\cos(2v)\sin(2\theta)}}{\sqrt{2}}. \end{aligned}$$

We can then bound these by $b_1 \leq \cos\left(\bar{\theta}\right)+\sin\left(\bar{\theta}\right)+1/\sqrt{2} \leq \sqrt{2}+1/\sqrt{2}$ and $b_2 \leq \cos\left(\bar{\theta}\right)+\sin\left(\bar{\theta}\right) \leq \sqrt{2}$, so that

$$|\partial_\lambda z| \leq \sqrt{2}(3/2+\alpha_\lambda) =: z_\lambda. \quad \text{(A.12)}$$

Note that $\partial_v[\cos(v-\lambda)\sin(v)] = \cos(\lambda-2v)$ and $\partial_v[\cos(v-\lambda)\cos(v)] = \sin(\lambda-2v)$. We can hence write the v derivative as

$$\partial_v z = a_1 + a_2 + a_3\partial_v\hat{\alpha}_0, \quad \text{where} \quad \text{(A.13)}$$

$$\begin{aligned} a_1 &= \cos(\lambda-2v)\sin(\hat{\alpha}_0)\cos\left(\bar{\theta}\right)+\sin(\lambda-2v)\cos(\hat{\alpha}_0)\sin\left(\bar{\theta}\right) \leq \cos\left(\bar{\theta}\right)+\sin\left(\bar{\theta}\right) \leq \sqrt{2} \\ a_2 &= -\frac{\cos(v-\lambda)\sqrt{1-\sin(2\theta)\cos(2v)}}{\sqrt{2}} - \frac{\sin(2\theta)\sin(2v)\sin(v-\lambda)}{\sqrt{2}\sqrt{1-\sin(2\theta)\cos(2v)}} \leq \sqrt{2} \\ a_3 &= \cos(v-\lambda)\left(\cos\left(\bar{\theta}\right)\sin(v)\cos(\hat{\alpha}_0)-\sin\left(\bar{\theta}\right)\cos(v)\sin(\hat{\alpha}_0)\right) \leq \cos\left(\bar{\theta}\right)+\sin\left(\bar{\theta}\right) \leq \sqrt{2}, \end{aligned}$$

and where we obtained the bound on a_2 using $|\frac{\sin(2v)\sin(2\theta)}{\sqrt{1-\sin(2\theta)\cos(2v))}}| \leq \sqrt{2}$. Hence, we can bound

$$|\partial_v z| \leq \sqrt{2}(2+\alpha_v) =: z_v. \quad \text{(A.14)}$$

Finally we compute the $\bar{\theta}$ derivative

$$\partial_{\bar{\theta}} z = c_1 + c_2 + c_3\partial_{\bar{\theta}}\hat{\alpha}_0 \quad \text{(A.15)}$$

where

$$\begin{aligned}
c_1 &= \cos(v-\lambda)\left(\cos(\hat{\alpha}_0)\cos(v)\cos\left(\bar{\theta}\right) - \sin(\hat{\alpha}_0)\sin(v)\sin\left(\bar{\theta}\right)\right) \leq \cos\left(\bar{\theta}\right) + \sin\left(\bar{\theta}\right) \leq \sqrt{2} \\
c_2 &= \frac{\cos(2\theta)\cos(2v)\sin(v-\lambda)}{\sqrt{2}\sqrt{1-\sin(2\theta)\cos(2v)}} \leq \frac{\cos(2\theta)}{\sqrt{2}\sqrt{1-\sin(2\theta)}} = \sqrt{\frac{1+\sin(2\theta)}{2}} \leq 1 \\
c_3 &= \cos(v-\lambda)\left(\cos(\hat{\alpha}_0)\sin(v)\cos\left(\bar{\theta}\right) - \sin(\hat{\alpha}_0)\cos(v)\sin\left(\bar{\theta}\right)\right) \leq \cos\left(\bar{\theta}\right) + \sin\left(\bar{\theta}\right) \leq \sqrt{2}
\end{aligned}$$

We hence obtain

$$|\partial_{\bar{\theta}} z| \leq \sqrt{2} + 1 + \sqrt{2}\alpha_{\bar{\theta}} =: z_\theta. \tag{A.16}$$

We now compute the derivatives of $\tilde{\epsilon}$:

$$\begin{aligned}
\partial_\lambda \tilde{\epsilon} &= -\partial_\lambda \hat{\alpha}_0(\cos(\theta)\sin(\hat{\alpha}_0+\lambda-2v) + \sin(\theta)\sin(\hat{\alpha}_0-\lambda+2v)) \\
&\quad + \sin(\theta)\sin(\hat{\alpha}_0-\lambda+2v) - \cos(\theta)\sin(\hat{\alpha}_0+\lambda-2v) \\
\partial_v \tilde{\epsilon} &= -\partial_v \hat{\alpha}_0(\cos(\theta)\sin(\hat{\alpha}_0+\lambda-2v) + \sin(\theta)\sin(\hat{\alpha}_0-\lambda+2v)) \\
&\quad -2\sin(\theta)\sin(\hat{\alpha}_0-\lambda+2v) + 2\cos(\theta)\sin(\hat{\alpha}_0+\lambda-2v) \\
\partial_\theta \tilde{\epsilon} &= -\partial_\theta \hat{\alpha}_0(\cos(\theta)\sin(\hat{\alpha}_0+\lambda-2v) + \sin(\theta)\sin(\hat{\alpha}_0-\lambda+2v)) + \cos(\theta)\cos(\hat{\alpha}_0-\lambda+2v) \\
&\quad - \sin(\theta)\cos(\hat{\alpha}_0+\lambda-2v).
\end{aligned}$$

Using the same techniques as above, we find the following bounds

$$\begin{aligned}
|\partial_\lambda \tilde{\epsilon}| &\leq \alpha_\lambda(\cos(\theta)+\sin(\theta)) + \cos(\theta) + \sin(\theta) \\
&\leq \sqrt{2}(\alpha_\lambda + 1) =: \epsilon_\lambda \\
|\partial_v \tilde{\epsilon}| &\leq \alpha_v(\cos(\theta)+\sin(\theta)) + 2\cos(\theta) + 2\sin(\theta) \\
&\leq \sqrt{2}(\alpha_v + 2) =: \epsilon_v \\
|\partial_\theta \tilde{\epsilon}| &\leq \alpha_\theta(\cos(\theta)+\sin(\theta)) + \cos(\theta) + \sin(\theta) \\
&\leq \sqrt{2}(\alpha_\theta + 1) =: \epsilon_\theta.
\end{aligned}$$

A.3 APPENDIX FOR H(AB|XYE)

A.3.1 General result on Grids

Suppose that we want to optimize the function:

$$\begin{aligned}
\min_{\mathbf{x}\in\mathcal{C},\mathbf{y}\in\mathcal{D}} \quad & g(\mathbf{x},\mathbf{y}) + h(\mathbf{x}) \\
\text{s.t.} \quad & \forall i: \quad f^i(\mathbf{x},\mathbf{y}) \geq 0.
\end{aligned} \tag{A.17}$$

where $\mathbf{x} = (x_0, x_1, \cdots, x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_0, y_1, \cdots, y_m) \in \mathbb{R}^m$. Also, we assume that f_i, g, h are differentiable functions in the respective domains.
Our goal is to eliminate certain parameters in the optimization problem, making it suitable for numerical optimizations. We further assume that there exists $\mathbf{x}_0$ such that

$$\inf_{\mathbf{x}\in\mathcal{C}} : h(\mathbf{x}) \geq h(\mathbf{x}_0). \tag{A.18}$$

Let $\Delta_{\mathbf{x}_0}$ be the difference vector defined by

$$\Delta_{\mathbf{x}_0}(\mathbf{x}) := \mathbf{x} - \mathbf{x}_0.$$

and the gradient restricted in $\mathbb{R}^n$ given by

$$\nabla_x g(\mathbf{x}, \mathbf{y}) = (\partial_{x_0} g(\mathbf{x}, \mathbf{y}), \partial_{x_1} g(\mathbf{x}, \mathbf{y}), \cdots, \partial_{x_n} g(\mathbf{x}, \mathbf{y}))$$

We can lower bound the objective function in terms of the function using the Taylor's theorem

$$\begin{aligned} g(\mathbf{x}, \mathbf{y}) + h(\mathbf{x}) &\geq g(\mathbf{x}_0, y) + h(\mathbf{x}_0) - \nabla_x g(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}).\Delta_{\mathbf{x}_0}(\tilde{\mathbf{x}}) \\ &\geq g(\mathbf{x}_0, y) + h(\mathbf{x}_0) - g_{\max}\Delta_{\max} \end{aligned} \tag{A.19}$$

for some $\tilde{\mathbf{x}} \in \mathcal{C}$, $\tilde{\mathbf{y}} \in \mathcal{D}$ and $g_{\max}$ and $\Delta_{\max}$ are any positive numbers satisfying:

$$g_{\max} \geq \max_{\mathbf{x}\in\mathcal{C}, \mathbf{y}\in\mathcal{D}} ||\nabla_x g(\mathbf{x}_0, \mathbf{y})|| \text{ and } \Delta_{\max} \geq \max_{\mathbf{x}\in\mathcal{C}} ||\Delta_{\mathbf{x}_0}||. \tag{A.20}$$

For the sake of completion, we also define the quantity

$$f^i_{\max} \geq \max_{\mathbf{x}\in\mathcal{C}, \mathbf{y}\in\mathcal{D}} ||\nabla_x f^i(\mathbf{x}_0, \mathbf{y})||. \tag{A.21}$$

Now consider the following result that shall enable us to compute the lower bound of the optimization problem (A.17):

Lemma 58. *Consider the optimization problem defined as follows:*

$$\begin{aligned} \min_{\mathbf{y}\in\mathcal{D}} \quad & g(\mathbf{x}_0, \mathbf{y}) + h(\mathbf{x}_0) - g_{\max}\Delta_{\max} \\ s.t. \quad \forall i : \quad & f^{\max}(\mathbf{x}_0, \mathbf{y}) \geq -f^i_{\max}\Delta_{\max}, \end{aligned} \tag{A.22}$$

where $\mathbf{x}_0$ is defined according to (A.18), and $g_{\max}$, $\Delta_{\max}$, and $f_{\max}$ are defined as in (A.20) and (A.21), respectively. This optimization problem serves as a lower bound on the problem (A.17).

Proof. In the discussion above, we showed that the objective function in (A.17) can be bounded from below using (A.19). It remains to show that the feasible set of the optimization problem (A.17) is subset of that of (A.22). This can also be shown using Taylor's Theorem. Let $(\mathbf{x}, \mathbf{y})$ be any point in the feasible set of (A.17), then

$$f^i(\mathbf{x}, \mathbf{y}) = f^i(\mathbf{x}_0, \mathbf{y}) - \nabla_x f^i(\tilde{\mathbf{x}_0}, \tilde{\mathbf{y}}).\Delta_{\mathbf{x}_0}(\tilde{\mathbf{x}})$$

for some $\tilde{\mathbf{x}_0} \in \mathcal{C}, \tilde{\mathbf{y}} \in \mathcal{D}$. Now, it is straightforward to see that,

$$f^i(\mathbf{x}_0, \mathbf{y}) - \nabla_x f^i(\tilde{\mathbf{x}_0}, \tilde{\mathbf{y}}).\Delta_{\mathbf{x}_0}(\tilde{\mathbf{x}}) \geq 0 \implies f^i(\mathbf{x}_0, \mathbf{y}) + f_{\max}\Delta_{\max} \geq 0.$$

Thus, the set of feasible points of the problem (A.17), remain feasible for the optimization problem (A.22) as well. □

A.4 USAGE OF EAT FOR DIFFERENT PROTOCOLS

A.4.1 PROTOCOL WITH RECYCLED INPUT RANDOMNESS (PROTOCOL 3)

For this protocol we want to extract randomness from the inputs and outputs. We hence set $C_i = A_i B_i X_i Y_i$ and take D_i to be trivial. When running a protocol, we do not generally know the set of EAT channels being used (these are set by the adversary), but instead only know that they have the no-signalling form, i.e., we have

$$\mathcal{M}(\rho_{A'B'}) = \sum_{abxy} |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes |u(a,b,x,y)\rangle\langle u(a,b,x,y)| \otimes \mathcal{M}^{abxy}(\rho_{A'B'}),$$

where $\mathcal{M}^{abxy}(\rho_{A'B'}) = p_{XY}(x,y)(\mathcal{E}^{x,a} \otimes \mathcal{F}^{y,b})(\rho_{A'B'})$ and $\{\mathcal{E}^{x,a}\}_a$ and $\{\mathcal{F}^{y,b}\}_b$ are instruments on A' and B' respectively (cf. (2.26)). Henceforth, the set $\mathfrak{G}$ will refer to all channels of this type.

In the CHSH protocol without spot-checking the rate function should be a lower bound on $H(ABXY|E) = 2 + H(AB|XYE)$ and we can form our rate function via $\mathrm{rate}(\{1-s, s\}) = 2 + F_{AB|XYE}(s)$ or $\mathrm{rate}(\{1-s, s\}) = 2 + F_{A|XYE}(s)$, the former being preferred as it is larger. A min-tradeoff function can then be obtained by taking the tangent at some point. Since $F_{AB|XYE}(s)$ is linear for $3/4 \leq s \leq \omega^*_{AB|XYE} \approx 0.847$, for experimentally relevant scores we can form the min-tradeoff function using the extension of this line to the domain $[0, 1]$, i.e., we can take $f(\{1-s, s\}) = 2 + G'_{AB|XYE}(\omega^*)(s - 3/4)$ in Theorem 1 when applying to Protocol 3, and in this case $d_C = d_A d_B d_X d_Y = 16$ and we get a bound on $H^{\epsilon_h}_{\min}(\mathbf{ABXY}|E)$. The theorem holds for all $\alpha \in (1, 2)$ and we can optimize over α to increase the bound.

A.4.2 SPOT-CHECKING CHSH PROTOCOL (PROTOCOL 1)

To use the EAT in the spot-checking CHSH protocol (Protocol 1) we set $C_i = A_iB_i$ and $D_i = X_iY_i$ in Theorem 1. The channels again have the no-signalling form mentioned above, and we can use either $F_{AB|00E}$ or $F_{A|00E}$ as the basis of our rate function. Since the two-sided version is larger, it is better to work with $F_{AB|00E}(s)$, and the related min-tradeoff function based on taking its tangent at some point. Modification is required to account for the spot-checking structure. If we let $g_t(\{1-s,s\})$ be the tangent of $F_{AB|00E}(s)$ taken at t then we can form the spot-checking min-tradeoff functions

$$f_t(\delta_u) = \begin{cases} \frac{1}{\gamma} g_t(\delta_u) + (1-\frac{1}{\gamma}) g_t(\delta_1) & u \in \{0,1\} \\ g_t(\delta_1) & u = \perp \end{cases}.$$

where t can be chosen (see e.g. [32, Section 5] for the argument behind this). Using this construction the following theorem can be derived (this is an adaptation of Theorem 3 in [31]).

Theorem 5 (Entropy Accumulation Theorem for spot-checking CHSH protocol)**.** *Let $\rho_{\mathbf{ABXYU}E}$ be a CQ state obtained using the spot-checking CHSH protocol (Protocol 1). Let Ω be the event $|\{i : U_i = 0\}| \leq n\gamma(1-\omega_{\exp}+\delta)$ with p_Ω being the probability of this event in $\rho_{\mathbf{ABXYU}E}$, and let $\rho_{\mathbf{ABXYU}E|\Omega}$ be the state conditioned on Ω. Let $\epsilon_h \in (0,1)$ and $\alpha \in (1,2)$. Then for any r such that $f_t(\mathrm{Freq}_{\mathbf{U}}) \geq r$ for all events in Ω we have*

$$\begin{aligned} H_{\min}^{\epsilon_h}(\mathbf{AB}|\mathbf{XY}E)_{\rho_{\mathbf{ABXY}E|\Omega}} > \quad & nr - \frac{\alpha}{\alpha-1}\log\left(\frac{1}{p_\Omega(1-\sqrt{1-\epsilon_h^2})}\right) \\ & + n \inf_{p \in \mathcal{Q}_{\mathfrak{G}}^{\gamma}} \left(\Delta(f_t,p) - (\alpha-1)V(f_t,p) - (\alpha-1)^2 K_\alpha(f_t)\right), \end{aligned}$$

where

$$\begin{aligned} \Delta(f_t,p) &:= F_{AB|XYE}(p(1)/\gamma) - f_t(p) \\ V(f_t,p) &= \frac{\ln 2}{2}\left(\log(9) + \sqrt{\mathrm{Var}_p(f_t)+2}\right)^2 \\ K_\alpha(f_t) &= \frac{1}{6\log(2-\alpha)^3 \ln 2} 2^{(\alpha-1)(2+\mathrm{Max}(f_t)-\mathrm{Min}_{\mathcal{Q}_{\mathfrak{G}}^{\gamma}}(f_t))} \ln^3(2^{2+\mathrm{Max}(f_t)-\mathrm{Min}_{\mathcal{Q}_{\mathfrak{G}}^{\gamma}}(f_t)} + \mathrm{e}^2). \end{aligned}$$

To use this theorem we can take $r = (F_{AB|00E}(t) + (\omega_{\exp} - \delta - t)F'_{AB|00E}(t))$ (cf. the discussion in [31]), and since the theorem holds for any t and α these can be optimized over.

A.4.3 PROTOCOL WITH BIASED LOCAL RANDOM NUMBERS (PROTOCOL 2)

To derive the randomness rates, we use Theorem 1 with $C_i = A_iB_i$ and $D_i = X_iY_i$, as in the previous subsection. What remains is to derive the min-tradeoff function and error terms. In this section, we compute these quantities and derive the expression for the completeness error in terms of the biasing parameters ζ_A, ζ_B and statistical error δ.

A.4.3.1 Deriving the min-tradeoff function

We seek a min-tradeoff function suitable for using with Protocol 2. To construct it we write the EAT channel in a slightly different way that is explicit in the input distribution p_{XY}:

$$\begin{aligned}\mathcal{M}_{p_{XY}}(\rho) \;=\; \sum_{abxy} p_{XY}(x,y)\,|a\rangle\langle a|_A \otimes |b\rangle\langle b|_B \otimes |x\rangle\langle x|_X \\ \otimes |y\rangle\langle y|_Y \otimes |(x,y,w)\rangle\langle(x,y,w)|_U \otimes \mathcal{M}^{x,y}_{a,b}(\rho)\,,\end{aligned} \tag{A.23}$$

where $\mathcal{M}^{x,y}_{a,b}$ are subnormalized channels. We can also consider the analogous channel where the U register only stores w (we use $\tilde{\mathcal{M}}$ to indicate this case). Next consider the entropy $H(AB|X=0,Y=0E)$, this entropy is calculated for the normalization of the state

$$(|0\rangle\langle 0|_X \otimes |0\rangle\langle 0|_Y \otimes \mathbb{1}_{ABUE})(\mathcal{M}_{p_{XY}} \otimes \mathcal{I}_E)(\rho_{RE})(|0\rangle\langle 0|_X \otimes |0\rangle\langle 0|_Y \otimes \mathbb{1}_{ABUE})\,.$$

For fixed $\{\mathcal{M}^{x,y}_{a,b}\}$, this is independent of p_{XY} (it is defined provided $p_{XY}(0,0) \neq 0$). We next note that for q as the distribution on the score (U) register

$$\begin{aligned}(\mathcal{M}_{p_{XY}} \otimes \mathcal{I}_E)(\rho_{RE})_U &= \sum_{abxyw} p_{XY}\,\mathrm{tr}\Big(\mathcal{M}^{x,y}_{a,b}(\rho_R)\Big)\,|(x,y,w)\rangle\langle(x,y,w)| \\ &= \sum_{xyw} q(x,y,w)\,|(x,y,w)\rangle\langle(x,y,w)| \\ (\mathcal{M}_{1/4} \otimes \mathcal{I}_E)(\rho_{RE})_U &= \sum_{abxyw} \frac{1}{4}\,\mathrm{tr}\Big(\mathcal{M}^{x,y}_{a,b}(\rho_R)\Big)\,|(x,y,w)\rangle\langle(x,y,w)| \\ &= \sum_{xyw} \frac{q(x,y,w)}{4p_{XY}}\,|(x,y,w)\rangle\langle(x,y,w)|\,,\end{aligned}$$

and hence

$$(\tilde{\mathcal{M}}_{1/4} \otimes \mathcal{I}_E)(\rho_{RE})_U = \sum_{xyw} \frac{q(x,y,w)}{4p_{XY}}\,|w\rangle\langle w|\,.$$

It follows that

$$\Big\{H(AB|X=0,Y=0,E)_{(\tilde{\mathcal{M}}_{1/4}\otimes\mathcal{I}_E)(\rho_{RE})} : (\tilde{\mathcal{M}}_{1/4}\otimes\mathcal{I}_E)(\rho_{RE})_U = (1-s)\,|0\rangle\langle 0| + s\,|1\rangle\langle 1|\,, \\ s=\sum_{xy}\frac{q(x,y,1)}{4p_{XY}}\Big\}$$

$$= \Big\{H(AB|X=0,Y=0,E)_{(\mathcal{M}_{p_{XY}}\otimes\mathcal{I}_E)(\rho_{RE})} : (\mathcal{M}_{p_{XY}}\otimes\mathcal{I}_E)(\rho_{RE})_U \\ = \sum_{xyw} q(x,y,w)\,|(x,y,w)\rangle\langle(x,y,w)|\Big\}.$$

Let $\mathfrak{G}_{\zeta^A,\zeta^B}$ be the set of channels for which X and Y are independent, X is 1 with probability ζ^A and Y is 1 with probability ζ^B.

Lemma 59. *The function $F_{AB|00E}$ as defined in the main text can be used to define a rate function for $\mathfrak{G}_{\zeta^A,\zeta^B}$ by taking $\mathrm{rate}_{\zeta^A,\zeta^B}(q) = F_{AB|00E}(\omega(q))$ for $q \in \mathcal{Q}_{\mathfrak{G}_{\zeta^A,\zeta^B}}$ where*

$$\omega(q) = \frac{1}{4}\sum_{xy}\frac{1}{p_X(x)p_Y(y)}q((x,y,1))\,. \tag{A.24}$$

Proof. We have

$$F_{AB|00E}(\omega(q)) := \inf_{(\tilde{\mathcal{M}},\rho_{RE})}\Big\{H(AB|X=0,Y=0,E)_{(\tilde{\mathcal{M}}_{1/4}\otimes\mathcal{I}_E)(\rho_{RE})} : (\tilde{\mathcal{M}}_{1/4}\otimes\mathcal{I}_E)(\rho_{RE})_U \\ = (1-\omega(q))\,|0\rangle\langle 0| + \omega(q)\,|1\rangle\langle 1|\Big\}$$

$$= \inf_{(\mathcal{M},\rho_{RE})}\Big\{H(AB|X=0,Y=0,E)_{(\mathcal{M}_{p_Xp_Y}\otimes\mathcal{I}_E)(\rho_{RE})} : (\mathcal{M}_{p_Xp_Y}\otimes\mathcal{I}_E)(\rho_{RE})_U \\ = \sum_{xyw} q(x,y,w)\,|(x,y,w)\rangle\langle(x,y,w)|\Big\}$$

$$\leq \inf_{(\mathcal{M},\rho_{RE})}\Big\{H(AB|XYE)_{(\mathcal{M}_{p_Xp_Y}\otimes\mathcal{I}_E)(\rho_{RE})} : (\mathcal{M}_{p_Xp_Y}\otimes\mathcal{I}_E)(\rho_{RE})_U \\ = \sum_{xyw} q(x,y,w)\,|(x,y,w)\rangle\langle(x,y,w)|\Big\},$$

and hence $F_{AB|00E}(\omega(q))$ is a rate function for $q \in \mathcal{Q}_{\mathfrak{G}_{\zeta^A,\zeta^B}}$. □

We can hence form min-tradeoff functions suitable for using with Protocol 2 by taking affine lower bounds to $F_{AB|00E}$. Taking the tangent to $F_{AB|00E}$ at t we have min-tradeoff function

$$f_t(q) := F_{AB|00E}(t) + F'_{AB|00E}(t)\left(\frac{1}{4}\sum_{x,y}\frac{1}{p_X(x)p_Y(y)}q((x,y,1)) - t\right),$$

or, in other words, considering deterministic distributions on $U = (x, y, w)$

$$f_t(\delta_{(x,y,w)}) = \begin{cases} \frac{F'_{AB|00E}(t)}{4p_X(x)p_Y(y)} + F_{AB|00E}(t) - tF'_{AB|00E}(t) & \text{if } w = 1 \\ F_{AB|00E}(t) - tF'_{AB|00E}(t) & \text{if } w = 0 \end{cases}$$

We have

$$\mathrm{Max}(f_t) = \frac{F'_{AB|00E}(t)}{4\zeta^A\zeta^B} + F_{AB|00E}(t) - tF'_{AB|00E}(t)$$

$$\mathrm{Min}_{\mathcal{Q}_{\mathfrak{S}_{\zeta^A,\zeta^B}}}(f_t) = F_{AB|00E}(t) - F'_{AB|00E}(t)\left(t - \frac{1}{2}\left(1 - \frac{1}{\sqrt{2}}\right)\right)$$

We now find a bound on $\mathrm{Var}_q(f_t)$ using the Bhatia-Davis bound [138].

Lemma 60 (Bhatia-Davis bound)**.** *Let X be a real-valued random variable with $\max(X) = M$, $\min(X) = m$ and $\mathbb{E}(X) = \mu$, then*

$$\mathrm{Var}_X \leq (M - \mu)(\mu - m)\,. \tag{A.25}$$

In our case, $M = \mathrm{Max}(f_t)$, $m = F_{AB|00E}(t) - tF'_{AB|00E}(t)$ and $\mu = \mathbb{E}_q(f_t) = F_{AB|00E}(t) + F'_{AB|00E}(t)\,(\omega(q) - t)$, where $\omega(q)$ is defined in (A.24). Thus,

$$\begin{aligned} \mathrm{Var}_q(f_t) &\leq (F'_{AB|00E}(t))^2\omega(q)\left(\frac{1}{4\zeta^A\zeta^B} - \omega(q)\right) \\ &\leq \begin{cases} (F'_{AB|00E}(t))^2\left(\frac{1}{4\zeta^A\zeta^B} - 1\right) & \text{if } \zeta^A\zeta^B < 1/8 \\ \left(\frac{F'_{AB|00E}(t)}{8\zeta^A\zeta^B}\right)^2 & \text{if } \zeta^A\zeta^B \geq 1/8 \end{cases} \end{aligned}$$

where we have optimized over $\omega(q) \in [0, 1]$ for the second inequality.

A.4.3.2 Completeness error

We can form a bound on the completeness error using Hoeffding's inequality [139].

Lemma 61 (Hoeffding's inequality)**.** *Let X_i be n i.i.d. random variables with $a \leq X_i \leq b$, $a, b \in \mathbb{R}$. If $S = \sum_i X_i$ and $\mu = \mathbb{E}(S)$. Then for $t > 0$*

$$\mathbb{P}(S - \mu \geq t) \leq \mathrm{e}^{-\frac{2t^2}{n(b-a)^2}}\,. \tag{A.26}$$

Theorem 6. *Suppose Protocol 2 is run using honest devices that behave in an i.i.d. fashion and that have an expected CHSH score ω_{exp}. The probability that the protocol aborts is no greater than*

$$\mathrm{e}^{-32n(\delta\zeta^A\zeta^B)^2}\,. \tag{A.27}$$

Proof. Recall the abort condition in the protocol, which states that $\omega < \omega_{\text{exp}} - \delta$ where

$$\omega = \frac{1}{4}\sum_{x,y}\frac{|\{i : U_i = (x,y,1)\}|}{np_X(x)p_Y(y)}\,.$$

We can write this as $\sum_i J_i$, where

$$J_i((x,y,w)) \;=\; \begin{cases} 0 & \text{if } w = 0 \\ 1/(4np_X(x)p_Y(y)) & \text{if } w = 1 \end{cases} \tag{A.28}$$

This construction gives $\mathbb{E}\left[\sum_i J_i\right] = n\mathbb{E}[J_i] = \sum_{xy}\frac{1}{4p_X(x)p_Y(y)}\mathbb{P}(U = (x,y,1))$. In an honest implementation of the protocol, the distribution on the register U takes the form

$$\mathbb{P}(U = (x,y,w)) = \begin{cases} p_X(x)p_Y(y)(1-\omega_{xy}) & \text{if } w = 0 \\ p_X(x)p_Y(y)\omega_{xy} & \text{if } w = 1 \end{cases} \tag{A.29}$$

where $\sum_{xy}\omega_{xy} = 4\omega_{\text{exp}}$, and hence $\mathbb{E}\left[\sum_i J_i\right] = \omega_{\text{exp}}$. The abort condition can be expressed as $\omega_{\text{exp}} - \sum_i J_i > \delta$. We have

$$\begin{aligned}\mathbb{P}(\omega_{\text{exp}} - \sum_i J_i > \delta) &= \mathbb{P}(\sum_i(-J_i) - (-\omega_{\text{exp}}) > \delta) \\ &\leq \mathrm{e}^{-32n(\delta\zeta^A\zeta^B)^2}\,,\end{aligned}$$

where we have used Hoeffding's inequality for the random variable $-J_i$ with $a = -1/(4n\zeta^A\zeta^B)$ and $b = 0$. □

A.4.4 ERROR PARAMETERS

Both Theorems 1 and 5 are stated in terms of the probability that the protocol does not abort, p_Ω, which is unknown to the users of the protocol. However, if we replace p_Ω by ϵ_{EAT}, then if $p_\Omega \geq \epsilon_{\text{EAT}}$ we have a correct bound on the entropy. On the other hand, if $p_\Omega < \epsilon_{\text{EAT}}$ then the protocol aborts with probability greater than $1 - \epsilon_{\text{EAT}}$. In other words, prior to running the protocol the probability that it will both not abort and that the entropy is not valid is at most ϵ_{EAT}. The soundness error of the protocol is $\epsilon_S = \max(\epsilon_{\text{EAT}}, 2\epsilon_h + \epsilon_{\text{EXT}})$, where ϵ_{EXT} is the extractor error (essentially the probability that the extraction fails). A summary of the aspects of extraction relevant to the present discussion and in the same notation as used here can be found in [31, Supplementary Information I C].

A.4.5 Application to $H(AB|E)$ and $H(A|E)$

Note that the EAT as stated in Theorem 1 cannot be directly used in conjunction with $H(AB|E)$ and $H(A|E)$. The basic reason is that the event Ω should be an event on $\mathbf{U}$, which in turn should be a deterministic function of $\mathbf{C}$ and $\mathbf{D}$. To use $H(AB|E)$ and $H(A|E)$ we need $\mathbf{D}$ to be empty and $\mathbf{C}$ to be $\mathbf{AB}$. This means the register $\mathbf{U}$ cannot depend on the inputs, $\mathbf{XY}$, but without a score that depends on the inputs we cannot certify non-classicality let alone randomness.
Since we do not have strong use cases for $H(AB|E)$ and $H(A|E)$, we do not consider possible extensions of the EAT in this work.
An alternative, which loses tightness, is to use an idea from [140, Appendix B.3]. Applying to the present case this would mean taking $\mathbf{D}$ to be empty and $\mathbf{C}$ to be either $\mathbf{ABV}$ or $\mathbf{AV}$, where V_i records whether the CHSH game was won on the ith round, with $U_i = V_i$. Then, proceeding with the former, because $H(ABV|E) \geq H(AB|E)$ we can base our min-tradeoff function on $H(AB|E)$, and we can use a chain rule to recover a bound on the smooth min entropy of $\mathbf{AB}$ given E from that of $\mathbf{ABV}$ given E. The bounds used in this approach are tightest when V has low entropy, so we expect better performance with spot-checking protocols.

A.5 DISCUSSION OF COMPOSABILITY

Throughout this work we consider a composable security definition, which means that, except with some small probability, the output randomness can be treated as a perfect random string as part of any larger protocol. Composable security definitions involve a distinguisher who tries to guess whether the real protocol or a hypothetical ideal protocol is being run. This distinguisher is allowed access to all the systems an eavesdropper has access to and is also assumed to learn whether or not the protocol was successful.
The main purpose of this appendix is to briefly discuss composability for protocols that recycle the input randomness. [The discussion here is not relevant for protocols without such recycling.] In general, input randomness (the strings $\mathbf{X}$ and $\mathbf{Y}$) is not directly reusable without processing [2]. For instance, the devices could be set up such that the protocol aborts unless $X_1 = 0$ and so if the protocol passes it is known that $X_1 = 0$. If $\mathbf{X}$ directly forms part of the output, then with probability $1/2$ one bit of the final output is known, which contradicts the security claim that the probability of distinguishing the output from perfect randomness is at most the soundness error.

Hence, in order to recycle the input randomness, it also has to undergo extraction to remove possible information that has leaked about it. The only mechanism for information leak allowed by the correct running of the protocol is whether it aborted or not, which constitutes at most one bit. This is easily remedied by compressing the concatenation of all inputs and outputs by one extra bit during the extraction step (the argument mirrors the more general case below).

One could also imagine more general protocols in which the length of the string output by the protocol is a variable (that depends on the actual score observed in the experimental run), in contrast to the protocols we use in the present paper whose outputs are all a pre-determined fixed length that depends on the expected CHSH score. Note that allowing varying output length requires a modified security definition with a different ideal state, as well as a modified analysis, both of which are beyond the scope of the present paper (see, e.g., [141, 142]). The main point we wish to make here is that if additional information L potentially leaks during the protocol then this could convey information about the input string. In protocols that recycle the input randomness, we can deal with this by additional compression in the extraction step. More precisely, Equation (3.21) of [143] implies $H_{\min}^{\epsilon_h}(R|SL) \geq H_{\min}^{\epsilon_h}(R|S) - \log d_L$, where d_L is the dimension of L.

In a protocol with variable output length, the distinguisher can get potentially useful information from the length L of the final random string. Given a bound on $H_{\min}^{\epsilon_h}(\mathbf{ABXY}|E)$ (to reiterate: obtaining this bound and the correct security definition are beyond the scope of this work), we can account for the information potentially conveyed by the final string length as follows. Conservatively, in a CHSH-based protocol with n rounds, the concatenation $\mathbf{ABXY}$ comprises at most $4n$ bits. Therefore, the length of the final output should satisfy

$$H_{\min}^{\epsilon_h}(\mathbf{ABXY}|EL) \geq H_{\min}^{\epsilon_h}(\mathbf{ABXY}|E) - \log d_L \geq H_{\min}^{\epsilon_h}(\mathbf{ABXY}|E) - \log(4n)\,.$$

Hence, if we reduce the length of the extractor output by $\log(4n)$ bits we can recycle the input randomness. Since the leading order term in $H_{\min}^{\epsilon_h}(\mathbf{ABXY}|E)$ is proportional to n, this reduction is minor.

Because we are working with device-independent protocols, the ongoing security of any randomness generated can be compromised if the devices used for one instance of the protocol are subsequently reused [60]. Hence, our discussion of security assumes devices are not reused (possible modifications to protocols that aim to allow restricted reuse are also discussed in [60]).

— B —

Appendix for semi-Device Independent protocols

B.1 USEFUL CLAIMS

This section is comprised of a series of standalone proofs that substantiate the claims presented in the main text. These proofs are independent of one another unless explicitly linked by a reference.

Lemma 62. *Let $\theta \in [0, \frac{\pi}{2}]$ and $x, h > 0$ are any reals (such that the functions below are defined), consider the functions*

$$\Lambda_0(x,h,\theta) := \Big(\Phi((x+h)\cos(\theta)) - \Phi(x+h)\Big) - \Big(\Phi(x\cos(\theta)) - \Phi(x)\Big) \quad \text{(B.1)}$$

$$\Lambda_1(x,h,\theta) := \Big(\Phi((x+h)\cos(\theta)) - \Phi(x\cos(\theta))\Big) \quad \text{(B.2)}$$

Then $\Lambda_0(x,h,\theta_1) \geq \Lambda(x,h,\theta_2)$ and $\Lambda_1(x,h,\theta_1) \geq \Lambda_1(x,h,\theta_2)$ if $\theta_1 > \theta_2$.

Proof.

$$\begin{aligned}\partial_\theta\Lambda_0(x,h,\theta) = \partial_\theta\Lambda_1(x,h,\theta) &= \Big(-\Phi'(x\cos(\theta)+h\cos(\theta))(1+h) + \Phi'(x\cos(\theta))\Big)\sin(\theta)\\ &\geq \Big(-\Phi'(x\cos(\theta)+h\cos(\theta)) + \Phi'(x\cos(\theta))\Big)\sin(\theta)\\ &\geq 0\end{aligned}$$

Where the last line follows from the fact that $-\Phi(x) > -\Phi(y)$ if $x > y$ and $x, y > 0$. □

Lemma 63. *Let $\theta \in [0, \frac{\pi}{2}]$. The function*

$$f(x) := \Phi(x\cos(\theta)) - \Phi(x) \quad \text{(B.3)}$$

is increasing for all $x \in [0,1]$.

Proof. Consider the function

$$\Lambda(x,h,\theta) := \Big(\Phi((x+h)\cos(\theta)) - \Phi(x+h)\Big) - \Big(\Phi(x\cos(\theta)) - \Phi(x)\Big) \tag{B.4}$$

We need to show that $\Lambda(x,h,\theta) > 0$. We first start by proving this claim for $\theta \leq \cos^{-1}\left(\frac{x}{x+h}\right)$. The choice of such θ shall be made clear in the second part of the proof. Suppose, that we are able to show that $\Lambda(x,h,\cos^{-1}\left(\frac{x}{x+h}\right)) \geq 0$, then we prove the $\Lambda(x,h,\theta) \geq 0$ for $\theta \geq \cos^{-1}\left(\frac{x}{x+h}\right)$ by using 62 and the obervation

$$\Lambda(x,h,\theta) \geq \Lambda(x,h,\cos^{-1}(\frac{x}{x+h})) \geq 0 \tag{B.5}$$

We now prove that $\Lambda(x,h,\theta) \geq 0$ for every $\theta \in [0,\cos^{-1}\left(\frac{x}{x+h}\right)]$. For any θ in this range, the sets $[x\cos(\theta), x]$ and $[(x+h)\cos(\theta), x+h]$ are disjoint. Using Taylor's theorem, we know that there exists $c_1 \in [x\cos(\theta), x]$ such that

$$\Phi(x\cos(\theta)) - \Phi(x) \;\; = \;\; -\Phi'(c_1)x(1-\cos(\theta))$$

Thus, $\exists c_2 \in [(x+h)\cos(\theta), x+h] > c_1$ such that

$$\begin{aligned}\Lambda(x,h,\theta) &= -\Phi'(c_2)(x+h)\cos(\theta) + \Phi'(c_1)x\cos(\theta) \\ &\geq \Big(-\Phi'(c_2) + \Phi'(c_1)\Big)x\cos(\theta) \\ &\geq 0 \end{aligned} \tag{B.6}$$

The last line here follows from the fact that $c_2 > c_1 > 0$ implies $-\Phi'(c_2) > -\Phi'(c_1)$. □

Lemma 64. *If* $\mathcal{H}_\lambda$ *is invariant under the action of* $\hat{O}$*. If* P_λ *is the projection onto the subspace* $\mathcal{H}_\lambda$ *then*

$$P_\lambda \hat{O} \in \mathcal{H}_\lambda \quad \text{and} \quad [P_\lambda, \hat{O}] = 0 \tag{B.7}$$

Proof. Let $|v\rangle \in \mathcal{H}_\lambda$. As $\mathcal{H}_\lambda$ is invariant under the action of $\hat{O}$, $\hat{O}|v\rangle = |w\rangle$ for some $|w\rangle \in \mathcal{H}_\lambda$. Then $P_\lambda|w\rangle \in \mathcal{H}_\lambda \implies P_\lambda\hat{O}|v\rangle \in \mathcal{H}_\lambda$.
Finally $P_\lambda\hat{O}|v\rangle = |w\rangle = \hat{O}|v\rangle = \hat{O}P_\lambda|v\rangle$. As $|v\rangle$ is any-arbitrary element of $\mathcal{H}_\lambda$ and $\forall\,|u\rangle \in \mathcal{H}: \quad P_\lambda\hat{O}|u\rangle \in \mathcal{H}_\lambda$, we must have $[P_\lambda, \hat{O}] = 0$. □

Another result that we use is the corollary of the no-signalling for our protocol

Lemma 65. *Let ρ^x_{AE} be any purification of ρ^x_A. Then, the von Neumann entropy of the reduced state on system E, denoted by $H(\rho^x_E)$, is equal to the von Neumann entropy of ρ^x_A, denoted by $H(\rho^x_A)$.*

Proof. First, note that any two purifications of ρ^x_A, ρ^x_{AE} and $\tilde{\rho}^x_{AE}$, are related by a unitary U acting on system E, i.e., $\tilde{\rho}^x_{AE} = (\mathbb{1}_A \otimes U)\rho^x_{AE}(\mathbb{1}_A \otimes U^\dagger)$. Using the properties of partial trace, we have

$$\begin{aligned}\mathrm{tr}_A(\tilde{\rho}^x_{AE}) &= \mathrm{tr}_A\left[(\mathbb{1}_A \otimes U)\rho^x_{AE}(\mathbb{1}_A \otimes U^\dagger)\right] \\ &= (\mathbb{1}_A \otimes U)\,\mathrm{tr}_A(\rho^x_{AE})(\mathbb{1}_A \otimes U^\dagger) \\ &= U\,\mathrm{tr}_A(\rho^x_{AE})U^\dagger.\end{aligned}$$

Since $H(U\rho U^\dagger) = H(\rho)$ holds for any state ρ and unitary U, it suffices to construct any specific purification of ρ^x_A and compute its entropy.
Let $|i\rangle$ be the eigenbasis of ρ^x_A, and let λ_i be the corresponding eigenvalues. Consider the following purification of ρ^x_A:

$$\rho^x_{AE} = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |i\rangle\langle j|_A \otimes |i\rangle\langle j|_E .$$

It is easy to verify that ρ^x_{AE} is indeed a purification of ρ^x_A. Moreover, it is easy to check that

$$\mathrm{tr}_A(\rho^x_{AE}) = \sum_i \lambda_i\, |i\rangle\langle i|_E .$$

Thus, we have

$$H(\rho^x_E) = H\left(\mathrm{tr}_A(\rho^x_{AE})\right) = H(\rho^x_A),$$

as desired. □

B.2 ELMINATING SOME PARAMETERS

The proofs below help in reduction of the parameter space for optimization problem for computing the rate functions.

Lemma 66. *Suppose we have the optimization problem:*

$$\begin{aligned}\inf \quad & f(x_1, x_2, \cdots, x_n, y) \\ s.t. \quad & \forall i : h_i(x_1, x_2, \cdots, x_n) \geq 0. \\ & g(x_1, x_2, \cdots, x_n, y) \geq 0\end{aligned} \tag{B.8}$$

The optimization problem B.8 is equivalent to the following problem:

$$\begin{aligned} \inf \quad & f(x_1, x_2, \cdots, x_n, y) \\ s.t. \quad & \forall i : h_i(x_1, x_2, \cdots, x_n) \geq 0 \\ & g(x_1, x_2, \cdots, x_n, y) \geq 0 \\ & y \in Y^* \end{aligned} \tag{B.9}$$

where Y^* is the following set:

$$Y^* := \{y \in \mathbb{R} : \quad \partial_y g(x_1, x_2, \cdots, x_n, y) = 0\}$$

Proof. We can solve the optimization problem (B.8) by writing the lagrangian:

$$\mathcal{L}(x_1, \cdots, x_n, y) := f(x_1, \cdots, x_n, y) - \sum_i \lambda_i h_i(x_1, \cdots, x_n) - \mu g(x_1, \cdots, x_n, y)$$

where $\lambda_i, \mu \in \mathbb{R}$ are some KKT multipliers. To determine the optimal parameter y, we must have that $\partial_y \mathcal{L} = 0$, which is equivalent to

$$\mu \partial_y g(x_1, \cdots, x_n, y) = 0.$$

Which gives the result unless $\mu = 0$. If $\mu = 0$, then the optimal solution for the (B.8) does not depend upon the value y, as y only appears in the constraint $g(x_1, \cdots, x_n, y) = 0$. Thus, we can choose $y \in Y^*$ without any loss of generality. □

B.3 MONOTONICITY OF RATES

This section consists of proofs for the monotonicity of the rate function $\mathcal{G}_{p_X}$.

Lemma 67. *The function $g_\omega(\Theta) := \mathcal{G}_{p_X}(\omega, \Theta)$ is increasing for all $\omega \in [\frac{1}{2}, 1]$ and $\Theta \geq \frac{1}{2}$.*

Proof. It suffices to show that the set of feasible points for the optimization problem $\mathcal{G}_{p_X}(\omega, \Theta)$ is a strict subset of the feasible set for the optimization problem for $\mathcal{G}_{p_X}(\omega, \Theta - \delta\Theta)$ for all $\delta\Theta \in [0, \Theta - \frac{1}{2}]$.
The parameter Θ appears in only one constraint denoted by:

$$\left(\sum_x \tilde{a}_x \cos(\xi_x)\right)^2 + \left(\sum_x \tilde{a}_x \sin(\xi_x)\right)^2 - \left(4\Theta - \sum_x \eta_x\right)^2 \geq 0. \tag{B.10}$$

Thus we aim to show that the set of points that obey

$$\left(\sum_x \tilde{a}_x \cos(\xi_x)\right)^2 + \left(\sum_x \tilde{a}_x \sin(\xi_x)\right)^2 - \left(4(\Theta - \delta\Theta) - \sum_x \eta_x\right)^2 \geq 0, \tag{B.11}$$

also obey (B.10). This can be easily proved by observing that:

$$-\left(4(\Theta - \delta\Theta) - \sum_x \eta_x\right)^2 = -\left(4\Theta - \sum_x \eta_x\right)^2 + 8\delta\Theta\left(\left(4\Theta - \sum_x \eta_x\right) - 2\delta\Theta\right).$$

Furthermore, it is easy to check that $\forall \delta\Theta \in [0, \Theta - \frac{1}{2}]$, the following holds:

$$8\delta\Theta\left(\left(4\Theta - \sum_x \eta_x\right) - 2\delta\Theta\right) \geq 0.$$

Thus, we can see that the set of points that obey (B.10), must automatically obey (B.11). □

Lemma 68. *The function* $g_\Theta(\omega) := \mathcal{G}_{p_X}(\omega, \Theta)$ *is increasing for all* $\Theta \in [\frac{1}{2}, 1]$ *and* $\omega \geq \frac{1}{2}$.

Proof. As in the previous proof, it suffices to show that the set of feasible points for the optimization problem $\mathcal{G}_{p_X}(\omega, \Theta)$ is a strict subset of the feasible set for the optimization problem for $\mathcal{G}_{p_X}(\omega - \delta\omega, \Theta)$ for all $\delta\omega \in [0, \Theta - \frac{1}{2}]$. This is even more straightforward in this case as

$$\sum_x \left(-\eta_x + (-1)^x \tilde{a}_x \cos(\xi_x)\right)) \geq 4\omega - 4 \geq 4(\omega - \delta\omega) - 4$$

holds for all $\delta\omega \in [0, 1]$. □

References

[1] Roger Colbeck. *Quantum and relativistic protocols for secure multi-party computation*. PhD thesis, University of Cambridge, 2007. Also available as arXiv:0911.3814.

[2] Roger Colbeck and Adrian Kent. Private randomness expansion with untrusted devices. *Journal of Physics A*, 44(9):095305, 2011.

[3] 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*, 464:1021–1024, 2010.

[4] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Ake Larsson, Carlos Abellan, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger. Significant-loophole-free test of Bells theorem with entangled photons. *Physical Review Letters*, 115:250401, 2015.

[5] 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*, 526:682–686, 2015.

[6] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, Scott Glancy, Deny R. Hamel, Michael S. Allman, Kevin J. Coakley, Shellee D. Dyer, Carson Hodge, Adriana E. Lita, Varun B. Verma, Camilla Lambrocco, Edward Tortorici, Alan L.

Migdall, Yanbao Zhang, Daniel R. Kumor, William H. Farr, Francesco Marsili, Matthew D. Shaw, Jeffrey A. Stern, Carlos Abellan, Waldimar Amaya, Valerio Pruneri, Thomas Jennewein, Morgan W. Mitchell, Paul G. Kwiat, Joshua C. Bienfang, Richard P. Mirin, Emanuel Knill, and Sae Woo Nam. Strong loophole-free test of local realism. *Physical Review Letters*, 115:250402, 2015.

[7] Yang Liu, Qi Zhao, Ming-Han Li, Jian-Yu Guan, Yanbao Zhang, Bing Bai, Weijun Zhang, Wen-Zhao Liu, Cheng Wu, Xiao Yuan, Hao Li, W. J. Munro, Zhen Wang, Lixing You, Jun Zhang, Xiongfeng Ma, Jingyun Fan, Qiang Zhang, and Jian-Wei Pan. Device-independent quantum random-number generation. *Nature*, 562:548–551, 2018.

[8] Ming-Han Li, Xingjian Zhang, Wen-Zhao Liu, Si-Ran Zhao, Bing Bai, Yang Liu, Qi Zhao, Yuxiang Peng, Jun Zhang, Yanbao Zhang, William J. Munro, Xiongfeng Ma, Qiang Zhang, Jingyun Fan, and Jian-Wei Pan. Experimental realization of device-independent quantum randomness expansion. *Physical Review Letters*, 126:050503, 2021.

[9] Thomas Van Himbeeck, Erik Woodhead, Nicolas J Cerf, Raúl García-Patrón, and Stefano Pironio. Semi-device-independent framework based on natural physical assumptions. *Quantum*, 1:33, 2017.

[10] Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nature Communications*, 2:411, 2011.

[11] Jon Barrett, Lucien Hardy, and Adrian Kent. No signalling and quantum key distribution. *Physical Review Letters*, 95:010503, 2005.

[12] Marco Tomamichel, Renato Renner, Christian Schaffner, and Adam Smith. Leftover hashing against quantum side information. In *Proceedings of the 2010 IEEE Symposium on Information Theory (ISIT10)*, pages 2703–2707, 2010.

[13] Yanbao Zhang, Honghao Fu, and Emanuel Knill. Efficient randomness certification by quantum probability estimation. *Physical Review Research*, 2:013016, 2020.

[14] Frédéric Dupuis, Omar Fawzi, and Renato Renner. Entropy accumulation. *Communications in Mathematical Physics*, 379:867–913, 2020.

[15] R. Arnon-Friedman, R. Renner, and T. Vidick. Simple and tight device-independent security proofs. *SIAM Journal on Computing*, 48:181–225, 2019.

[16] Richard Kueng. Convex optimization in quantum information theory. Lecture notes for ICMAT term in quantum information theory, 2023. The notes are available on

the following link at the time of writing this thesis: https://www.icmat.es/RT/2023/QIT/download/week1/kueng-optimization-notes.pdf.

[17] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019.

[18] Steven Diamond and Stephen Boyd. Cvxpy: A python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[19] Carl W Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969.

[20] Alexander Semenovich Holevo. Optimal quantum measurements. *Teoreticheskaya i Matematicheskaya Fizika*, 17(3):319–326, 1973.

[21] Hamza Fawzi. Semidefinite programming in quantum information. Advanced School on Optimization Methods in Quantum Information, 2023. Lecture series at ICMAT. See https://www.damtp.cam.ac.uk/user/hf323/qip2021_tutorial.html for slides.

[22] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas, editors. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.

[23] Peter Wittek. Algorithm 950: Ncpol2sdpasparse semidefinite programming relaxations for polynomial optimization problems of noncommuting variables. *ACM Transactions on Mathematical Software (TOMS)*, 41(3):1–12, 2015.

[24] Guillaume Sagnol and Maximilian Stahlberg. PICOS: A Python interface to conic optimization solvers. *Journal of Open Source Software*, 7(70):3915, February 2022.

[25] Robert König, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, 2009.

[26] Anindya De, Christopher Portmann, Thomas Vidick, and Renato Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41:915–940, 2012.

[27] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[28] Andrzej Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.

[29] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009.

[30] Tony Metger, Omar Fawzi, David Sutter, and Renato Renner. Generalised entropy accumulation. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 844–850. IEEE, 2022.

[31] Wen-Zhao Liu, Ming-Han Li, Sammy Ragy, Si-Ran Zhao, Bing Bai, Yang Liu, Peter J. Brown, Jun Zhang, Roger Colbeck, Jingyun Fan, Qiang Zhang, and Jian-Wei Pan. Device-independent randomness expansion against quantum side information. *Nature Physics*, 17:448–451, 2021.

[32] Frederic Dupuis and Omar Fawzi. Entropy accumulation with improved second-order term. *IEEE Transactions on Information Theory*, 65:7596–7612, 2019.

[33] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, pages 503–509, Los Alamitos, CA, USA, 1998. IEEE Computer Society.

[34] Lynden K Shalm, Yanbao Zhang, Joshua C Bienfang, Collin Schlager, Martin J Stevens, Michael D Mazurek, Carlos Abellán, Waldimar Amaya, Morgan W Mitchell, Mohammad A Alhejji, Honghao Fu, Joel Ornstein, Richard P. Mirin, Sae Woo Nam, and Emanuel Knill. Device-independent randomness expansion with entangled photons. *Nature Physics*, 17:452–456, 2021.

[35] Peter Bierhorst, Emanuel Knill, Scott Glancy, Yanbao Zhang, Alan Mink, Stephen Jordan, Andrea Rommal, Yi-Kai Liu, Bradley Christensen, Sae Woo Nam, Martin J. Stevens, and Lynden K. Shalm. Experimentally generated randomness certified by the impossibility of superluminal signals. *Nature*, 556:223–227, 2018.

[36] Umesh Vazirani and Thomas Vidick. Certifiable quantum dice or, testable exponential randomness expansion. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, STOC '12, pages 61–76, 2012.

[37] Carl A. Miller and Yaoyun 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, pages 417–426, New York, NY, USA, 2014. ACM.

[38] Carl A. Miller and Yaoyun Shi. Universal security for randomness expansion from the spot-checking protocol. *Siam Journal of Computing*, 46:1304–1335, 2017.

[39] William Slofstra. The set of quantum correlations is not closed. *Forum of Mathematics, Pi*, 7:e1, 2019.

[40] Károly F. Pál and Tamás Vértesi. Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems. *Physical Review A*, 82:022116, 2010.

[41] C. Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la S. M. F.*, 3:103–174, 1875.

[42] Stefano Pironio, Antonio Acin, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009.

[43] Peter J. Brown, Hamza Fawzi, and Omar Fawzi. Device-independent lower bounds on the conditional von Neumann entropy. e-print arXiv:2106.13692, 2021.

[44] Antonio Acín, Serge Massar, and Stefano Pironio. Randomness versus nonlocality and entanglement. *Physical Review Letters*, 108(10):100402, 2012.

[45] Erik Woodhead, Antonio Acín, and Stefano Pironio. Device-independent quantum key distribution with asymmetric CHSH inequalities. *Quantum*, 5:443, 2021.

[46] Lewis Wooltorton, Peter Brown, and Roger Colbeck. Tight analytic bound on the trade-off between device-independent randomness and nonlocality. *Physical Review Letters*, 129(15):150403, 2022.

[47] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419, 2014.

[48] Matthew Coudron and Henry Yuen. Infinite randomness expansion with a constant number of devices. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 427–436, 2014.

[49] Roger Colbeck and Renato Renner. Free randomness can be amplified. *Nature Physics*, 8:450–454, 2012.

[50] Stefano Pironio and Serge Massar. Security of practical private randomness generation. *Physical Review A*, 87:012336, 2013.

[51] Peter J. Brown, Sammy Ragy, and Roger Colbeck. A framework for quantum-secure device-independent randomness expansion. *IEEE Transactions on Information Theory*, 66:2964–2987, 2020.

[52] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Physical Review Letters*, 98(1):010401, 2007.

[53] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[54] Pavel Sekatski, Jean-Daniel Bancal, Xavier Valcarce, Ernest Y-Z Tan, Renato Renner, and Nicolas Sangouard. Device-independent quantum key distribution from generalized CHSH inequalities. *Quantum*, 5:444, 2021.

[55] Ernest Y-Z Tan, Pavel Sekatski, Jean-Daniel Bancal, René Schwonnek, Renato Renner, Nicolas Sangouard, and Charles C-W Lim. Improved DIQKD protocols with finite-size analysis. *Quantum*, 6:880, 2022.

[56] Ernest Y-Z Tan, René Schwonnek, Koon Tong Goh, Ignatius William Primaatmaja, and Charles C-W Lim. Computing secure key rates for quantum cryptography with untrusted devices. *npj Quantum Information*, 7(1):158, 2021.

[57] Peter J. Brown, Hamza Fawzi, and Omar Fawzi. Computing conditional entropies for quantum correlations. *Nature Communications*, 12:575, 2021.

[58] Florian J Curchod, Markus Johansson, Remigiusz Augusiak, Matty J Hoban, Peter Wittek, and Antonio Acín. Unbounded randomness certification using sequences of measurements. *Physical Review A*, 95(2):020102, 2017.

[59] Samuel L. Braunstein and Carlton M. Caves. Wringing out better Bell inequalities. *Annals of Physics*, 202(1):22–56, 1990.

[60] Jonathan Barrett, Roger Colbeck, and Adrian Kent. Memory attacks on device-independent quantum cryptography. *Physical Review Letters*, 106:010503, 2013.

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

[62] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

[63] Te Sun Hao and Mamoru Hoshi. Interval algorithm for random number generation. *IEEE Transactions on Information Theory*, 43:599–611, 1997.

[64] Hamid Tebyanian, Marco Avesani, Giuseppe Vallone, and Paolo Villoresi. Semi-device-independent randomness from d-outcome continuous-variable detection. *Physical Review A*, 104:062424, 2021.

[65] Jonatan Bohr Brask, Anthony Martin, William Esposito, Raphael Houlmann, Joseph Bowles, Hugo Zbinden, and Nicolas Brunner. Megahertz-rate semi-device-independent quantum random number generators based on unambiguous state discrimination. *Physical Review Applied*, 7:054018, 2017.

[66] Davide Rusca, Hamid Tebyanian, Anthony Martin, and Hugo Zbinden. Fast self-testing quantum random number generator based on homodyne detection. *Applied Physics Letters*, 116(26), 2020.

[67] Marco Avesani, Hamid Tebyanian, Paolo Villoresi, and Giuseppe Vallone. Semi-device-independent heterodyne-based quantum random-number generator. *Physical Review Applied*, 15:034034, 2021.

[68] Hamid Tebyanian, Mujtaba Zahidy, Marco Avesani, Andrea Stanco, Paolo Villoresi, and Giuseppe Vallone. Semi-device independent randomness generation based on quantum state's indistinguishability. *Quantum Science and Technology*, 6(4):045026, 2021.

[69] Thomas Van Himbeeck and Stefano Pironio. Correlations and randomness generation based on energy constraints. *arXiv preprint arXiv:1905.09117*, 2019.

[70] André Stefanov, Nicolas Gisin, Olivier Guinnard, Laurent Guinnard, and Hugo Zbinden. Optical quantum random number generator. *Journal of Modern Optics*, 47(4):595–598, 2000.

[71] Thomas Jennewein, Ulrich Achleitner, Gregor Weihs, Harald Weinfurter, and Anton Zeilinger. A fast and compact quantum random number generator. *Review of Scientific Instruments*, 71(4):1675–1680, 2000.

[72] Xiongfeng Ma, Xiao Yuan, Zhu Cao, Bing Qi, and Zhen Zhang. Quantum random number generation. *npj Quantum Information*, 2(1):1–9, 2016.

[73] Marco Avesani, Davide G Marangon, Giuseppe Vallone, and Paolo Villoresi. Source-device-independent heterodyne-based quantum random number generator at 17 gbps. *Nature communications*, 9(1):5365, 2018.

[74] Davide G Marangon, Giuseppe Vallone, and Paolo Villoresi. Source-device-independent ultrafast quantum random number generation. *Physical Review Letters*, 118(6):060503, 2017.

[75] Zhu Cao, Hongyi Zhou, and Xiongfeng Ma. Loss-tolerant measurement-device-independent quantum random number generation. *New Journal of Physics*, 17(12):125011, 2015.

[76] You-Qi Nie, Jian-Yu Guan, Hongyi Zhou, Qiang Zhang, Xiongfeng Ma, Jun Zhang, and Jian-Wei Pan. Experimental measurement-device-independent quantum random-number generation. *Physical Review A*, 94(6):060301, 2016.

[77] Stefano Pirandola, Carlo Ottaviani, Gaetana Spedalieri, Christian Weedbrook, Samuel L Braunstein, Seth Lloyd, Tobias Gehring, Christian S Jacobsen, and Ulrik L Andersen. High-rate measurement-device-independent quantum cryptography. *Nature Photonics*, 9(6):397–402, 2015.

[78] Yeong-Cherng Liang, Tamás Vértesi, and Nicolas Brunner. Semi-device-independent bounds on entanglement. *Physical Review A*, 83(2):022108, 2011.

[79] Marcin Pawowski and Nicolas Brunner. Semi-device-independent security of one-way quantum key distribution. *Physical Review A*, 84(1):010302, 2011.

[80] Hong-Wei Li, Zhen-Qiang Yin, Yu-Chun Wu, Xu-Bo Zou, Shuang Wang, Wei Chen, Guang-Can Guo, and Zheng-Fu Han. Semi-device-independent random-number expansion without entanglement. *Physical Review A*, 84(3):034301, 2011.

[81] Nicolas Brunner, Miguel Navascués, and Tamás Vértesi. Dimension witnesses and quantum state discrimination. *Physical Review Letters*, 110(15):150501, 2013.

[82] Joseph Bowles, Marco Túlio Quintino, and Nicolas Brunner. Certifying the dimension of classical and quantum systems in a prepare-and-measure scenario with independent devices. *Physical Review Letters*, 112(14):140407, 2014.

[83] David Drahi, Nathan Walk, Matty J Hoban, Aleksey K Fedorov, Roman Shakhovoy, Akky Feimov, Yury Kurochkin, W Steven Kolthammer, Joshua Nunn, Jonathan Barrett, et al. Certified quantum random numbers from untrusted light. *Physical Review X*, 10(4):041048, 2020.

[84] Armin Tavakoli. Semi-device-independent framework based on restricted distrust in prepare-and-measure experiments. *Physical Review Letters*, 126(21):210503, 2021.

[85] Hong-Wei Li, Marcin Pawłowski, Zhen-Qiang Yin, Guang-Can Guo, and Zheng-Fu Han. Semi-device-independent randomness certification using n 1 quantum random access codes. *Physical Review A*, 85(5):052308, 2012.

[86] Anubhav Chaturvedi, Máté Farkas, and Victoria J Wright. Characterising and bounding the set of quantum behaviours in contextuality scenarios. *Quantum*, 5:484, 2021.

[87] Kieran Flatt, Hanwool Lee, Carles Roch I Carceller, Jonatan Bohr Brask, and Joonwoo Bae. Contextual advantages and certification for maximum-confidence discrimination. *PRX Quantum*, 3(3):030337, 2022.

[88] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[89] Yves Lucet. Faster than the fast legendre transform, the linear-time legendre transform. *Numerical Algorithms*, 16:171–185, 1997.

[90] Lorenzo Contento, Alexandre Ern, and Rossana Vermiglio. A linear-time approximate convex envelope algorithm using the double legendre–fenchel transform with application to phase separation. *Computational Optimization and Applications*, 60:231–261, 2015.

[91] Davide Rusca, Thomas Van Himbeeck, Anthony Martin, Jonatan Bohr Brask, Weixu Shi, Stefano Pironio, Nicolas Brunner, and Hugo Zbinden. Self-testing quantum random-number generator based on an energy bound. *Physical Review A*, 100(6):062338, 2019.

[92] J. F. Clauser and M. A. Horne. Experimental consequences of objective local theories. *Physical Review D*, 10(2):526–535, 1974.

[93] Jonathan Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75:032304, 2007.

[94] Martin Plávala. General probabilistic theories: An introduction. *arXiv preprint arXiv:2103.07469*, 2021.

[95] David Gross, Markus Müller, Roger Colbeck, and Oscar CO Dahlsten. All reversible dynamics in maximally nonlocal theories are trivial. *Physical Review Letters*, 104(8):080402, 2010.

[96] Sandu Popescu and Daniel Rohrlich. Which states violate Bell's inequality maximally? *Physics Letters A*, 169(6):411–414, 1992.

[97] Ll Masanes, Antonio Acín, and Nicolas Gisin. General properties of nonsignaling theories. *Physical Review A*, 73(1):012112, 2006.

[98] Wim Van Dam. Implausible consequences of superstrong nonlocality. *Natural Computing*, 12:9–12, 2013.

[99] Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Physical Review Letters*, 96(25):250401, 2006.

[100] Noah Linden, Sandu Popescu, Anthony J Short, and Andreas Winter. Quantum nonlocality and beyond: limits from nonlocal computation. *Physical Review Letters*, 99(18):180502, 2007.

[101] Anthony J Short, Sandu Popescu, and Nicolas Gisin. Entanglement swapping for generalized nonlocal correlations. *Physical Review A*, 73(1):012101, 2006.

[102] Charles H Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Physical Review Letters*, 70(13):1895, 1993.

[103] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[104] Lucien Hardy. Quantum theory from five reasonable axioms. e-print quant-ph/0101012, 2001.

[105] Giulio Chiribella, Giacomo Mauro DAriano, and Paolo Perinotti. Informational derivation of quantum theory. *Physical Review A*, 84:012311, 2011.

[106] Lluís Masanes and Markus P. Müller. A derivation of quantum theory from physical requirements. *New Journal of Physics*, 13:063001, 2011.

[107] Markus P Müller and Cozmin Ududec. Structure of reversible computation determines the self-duality of quantum theory. *Physical Review Letters*, 108(13):130401, 2012.

[108] Howard Barnum and Alexander Wilce. Information processing in convex operational theories. *Electronic Notes in Theoretical Computer Science*, 270(1):3–15, 2011.

[109] Howard Barnum, Ciarán M Lee, and John H Selby. Oracles and query lower bounds in generalised probabilistic theories. *Foundations of Physics*, 48:954–981, 2018.

[110] Andrew JP Garner. Interferometric computation beyond quantum theory. *Foundations of Physics*, 48(8):886–909, 2018.

[111] Marius Krumm and Markus P Müller. Quantum computation is the unique reversible circuit model for which bits are balls. *npj Quantum Information*, 5(1):7, 2019.

[112] Jonathan Barrett, Niel de Beaudrap, Matty J Hoban, and Ciarán M Lee. The computational landscape of general physical theories. *npj Quantum Information*, 5(1):41, 2019.

[113] Howard Barnum, Oscar CO Dahlsten, Matthew Leifer, and Ben Toner. Nonclassicality without entanglement enables bit commitment. In *2008 IEEE Information Theory Workshop*, pages 386–390. IEEE, 2008.

[114] Giorgos Eftaxias, Mirjam Weilenmann, and Roger Colbeck. Advantages of multicopy nonlocality distillation and its application to minimizing communication complexity. *Physical Review Letters*, 130(10):100201, 2023.

[115] Howard Barnum, Carl Philipp Gaebler, and Alexander Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories. *Foundations of Physics*, 43:1411–1427, 2013.

[116] Paul Skrzypczyk and Nicolas Brunner. Couplers for non-locality swapping. *New Journal of Physics*, 11(7):073014, 2009.

[117] Howard Barnum, Carl Philipp Gaebler, and Alexander Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories. *Foundations of Physics*, 43:1411–1427, 2013.

[118] Martin Plávala. Conditions for the compatibility of channels in general probabilistic theory and their connection to steering and Bell nonlocality. *Physical Review A*, 96(5):052127, 2017.

[119] Manik Banik. Measurement incompatibility and schrödinger-einstein-podolsky-rosen steering in a class of probabilistic theories. *Journal of Mathematical Physics*, 56(5), 2015.

[120] Robert W Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110, 2007.

[121] David Schmid, John H Selby, Elie Wolfe, Ravi Kunjwal, and Robert W Spekkens. Characterization of noncontextuality in the framework of generalized probabilistic theories. *PRX Quantum*, 2(1):010331, 2021.

[122] David Schmid, John H Selby, Matthew F Pusey, and Robert W Spekkens. A structure theorem for generalized-noncontextual ontological models. *arXiv preprint arXiv:2005.07161*, 2020.

[123] Mirjam Weilenmann and Roger Colbeck. Analysing causal structures in generalised probabilistic theories. *Quantum*, 4:236, 2020.

[124] Giulio Chiribella and Carlo Maria Scandolo. Entanglement and thermodynamics in general probabilistic theories. *New Journal of Physics*, 17(10):103027, 2015.

[125] Carlo Maria Scandolo. Information-theoretic foundations of thermodynamics in general probabilistic theories. *arXiv preprint arXiv:1901.08054*, 2019.

[126] Giulio Chiribella and Carlo Maria Scandolo. Microcanonical thermodynamics in general physical theories. *New Journal of Physics*, 19(12):123043, 2017.

[127] Venkatesh Vilasini, Nuriya Nurgalieva, and Lídia del Rio. Multi-agent paradoxes beyond quantum theory. *New Journal of Physics*, 21(11):113028, 2019.

[128] Nick Ormrod, V Vilasini, and Jonathan Barrett. Which theories have a measurement problem? *arXiv preprint arXiv:2303.03353*, 2023.

[129] Victoria J Wright and Stefan Weigert. General probabilistic theories with a gleason-type theorem. *Quantum*, 5:588, 2021.

[130] Vincenzo Fiorentino and Stefan Weigert. A quantum theory with non-collapsing measurements. *arXiv preprint arXiv:2303.13411*, 2023.

[131] Peter Janotta and Raymond Lal. Generalized probabilistic theories without the no-restriction hypothesis. *Physical Review A*, 87(5):052131, 2013.

[132] David Schmid, John H Selby, Matthew F Pusey, and Robert W Spekkens. A structure theorem for generalized-noncontextual ontological models. *arXiv preprint arXiv:2005.07161*, 2020.

[133] Howard Barnum, Ciarán M Lee, Carlo Maria Scandolo, and John H Selby. Ruling out higher-order interference from purity principles. *Entropy*, 19(6):253, 2017.

[134] Bob Coecke and Ross Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13(4):043016, 2011.

[135] Bob Coecke, Ross Duncan, Aleks Kissinger, and Quanlong Wang. Generalised compositional theories and diagrammatic reasoning. *Quantum Theory: Informational Foundations and Foils*, pages 309–366, 2016.

[136] Bob Coecke. Zx-lectures, 2023.

[137] Paul Skrzypczyk, Nicolas Brunner, and Sandu Popescu. Emergence of quantum correlations from nonlocality swapping. *Physical Review Letters*, 102(11):110402, 2009.

[138] Rajendra Bhatia and Chandler Davis. A better bound on the variance. *The American Mathematical Monthly*, 107:353–357, 2000.

[139] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[140] 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*, 4:035011, 2019.

[141] Christopher Portmann and Renato Renner. Security in quantum cryptography. *Reviews of Modern Physics*, 94:025008, 2022.

[142] M. Ben-Or, Michal Horodecki, D. W. Leung, D. Mayers, and J. Oppenheim. The universal composable security of quantum key distribution. In *Second Theory of Cryptography Conference*, volume 3378, pages 386–406. Springer, 2005.

[143] Renato Renner. *Security of quantum key distribution*. PhD thesis, Swiss Federal Institute of Technology, Zurich, 2005. Also available as quant-ph/0512258.