ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF PHYSICS AND ASTRONOMY "A. RIGHI"

SECOND CYCLE DEGREE

PHYSICS

An input-randomness-free device-independent cryptographic protocol for quantum networks

Supervisor

Prof. Lorenzo Piroli

Defended by

Lorenzo Caramelli

Co-supervisors

Prof. Antonio Acín,
Dr. Tamás Kriváchy,
Dr. Raja Yehia

Graduation Session 12/2025

Academic Year 2024/2025

Information is physical.

Rolf Landauer

Abstract

Mentre la nostra comprensione della teoria dell'informazione quantistica si espande, è ancora da verificare che questa crescita sia accompagnata dallo sviluppo di applicazioni pratiche. I network quantistici e la crittografia quantistica sono due rami promettenti da questo punto di vista e sono ogni giorno più intrecciati. In particolare, un fenomeno affascinante è la presenza nonlocalità di Bell senza bisogno di diversi sistemi di misura, a patto che le sorgenti della rete siano indipendenti, i.e. in ogni round ciascuna sorgente produce lo stesso stato e ciascun partecipante effettua la stessa misura. Questa possibilità è in netto contrasto con il caso di due soli utenti, dove per poter certificare la presenza di nonlocalità, e quindi di randomness in uscita, è necessario che i partecipanti abbiano a disposizione delle sorgenti di randomness per scegliere casualmente quale misura effettuare. Partendo dalla nonlocalità senza randomness nei network, proponiamo un protocollo di device-independent quantum key distribution (che può essere esteso ad un protocollo di quantum conference key agreement se tra più utenti) che permette la creazione di una chiave segreta senza aver bisogno di randomness iniziale. Per fare ciò partiamo da risultati recenti nei campi della nonlocalità nei network e della sua robustezza, mostrando come in un anello di N sorgenti bipartite, bastino alcune assunzioni comuni per poter eseguire il protocollo, mentre nessuna fiducia è richiesta dai sistemi di misura. Successivamente diamo una dimostrazione device-independent della sicurezza del protocollo date queste assunzioni in un sistema IID. Riassumendo proponiamo un protocollo di Device-Independent Quantum Key Distribution che è il primo del suo genere, nativo dei networks quantistici e che non richiede randomness in entrata dagli utenti per produrre randomness in uscita.

Abstract

As our foundational understanding of quantum information theory broadens, it remains to be seen whether novel foundational developments will lead to corresponding applications. Promising branches are quantum networks and quantum cryptography, which are more intertwined by the day. In particular, a fascinating phenomena is that in networks with independent sources, Bell nonlocality can take place without the need for measurement settings for the parties, i.e. in each round, each source is asked to produce the same state and each party measures the same measurement. This is in stark contrast to standard Bell nonlocality, where seed randomness is necessary to certify nonlocality and through it, randomness of the outcomes. Building on nonlocality without seed randomness in networks, we propose a device-independent quantum key distribution protocol (which can be extended to quantum conference key agreement if it involves more than 2 parties) which allows for the generation of shared secret keys without the need for input randomness. We rely on recent results on nonlocality in networks and its robustness, showing how in a ring network with N bipartite sources, one must only make standard assumptions in order to run the protocol, while no trust is required from the devices themselves. We thus give a device-independent proof of security under this network assumption in the IID setting. In summary, we propose a first-of-a-kind protocol for Device-Independent Quantum Key Distribution which is native to the network setting, and which does not demand input randomness from the parties in order to produce output randomness.

Contents

Abstract (italian) I
Abstract (english) II

Introduction **1**

1 Quantum Information Theory **4**
1.1 Qubits 4
1.1.1 Multiple Qubits 6
1.1.2 Entanglement and Bell States 6
1.1.3 Density Operator 7
1.1.4 Fidelity and Trace Distance 8
1.2 Measurements 9
1.3 The EPR paradox and Nonlocality 11
1.3.1 Bell Inequalities 12
1.3.2 The Bell theorem and the CHSH inequality 12
1.3.3 Polytopes 15
1.4 Entropy 19
1.4.1 Classical Entropies 20
1.4.1.1 Mathematical Properties 21
1.4.1.2 Joint and Conditional Entropy 23
1.4.2 Quantum Entropies 26
1.4.2.1 Mathematical Properties 27
1.4.2.2 Joint and Conditional Entropy 29
1.4.2.3 (Smooth) Min- and (Smooth) Max-Entropy 30
1.4.3 Entropy relations 33
1.4.3.1 Chain Rules 33
1.4.3.2 Entropy Accumulation Theorem 35
1.4.3.3 Quantum Asymptotic Equipartition Property . . . 37

2 Quantum Cryptography **39**
2.1 Quantum Key Distribution 40
2.1.1 Prepare-and-Measure and Entanglement-based protocols . . 40

2.1.2 Quantum hacking . . . 43
2.1.3 Device-Independent Quantum Key Distribution . . . 44
2.1.4 Composable security . . . 45
2.2 Classical post-processing . . . 46
2.2.1 Information Reconciliation . . . 46
2.2.2 Privacy Amplification . . . 47
2.3 Quantum Conference Key Agreement . . . 49
2.3.1 Composable Security . . . 51
2.3.2 Device-Independent Conference Key Agreement . . . 51

3 Quantum Networks **53**
3.1 Overview of quantum networks . . . 53
3.1.1 Polytopes for networks . . . 55
3.1.2 Genuine Network Nonlocality (GNNL) . . . 56
3.1.3 Self-Testing and Token-Counting . . . 59
3.2 Triangle Network . . . 61
3.2.1 The RGB4 Distribution . . . 61
3.2.2 Inflation . . . 63

4 Cryptographic Protocol **68**
4.1 Protocol for A and B . . . 68
4.1.1 Setup . . . 68
4.1.2 Verification . . . 69
4.1.3 Key generation . . . 70
4.1.4 Raw key privacy amplification . . . 71
4.2 QCKA extension . . . 72
4.2.1 Setup and verification . . . 72
4.2.2 Key generation and privacy amplification . . . 73
4.2.2.1 $N = 4$ example . . . 74
4.3 Security . . . 75
4.3.1 Assumptions . . . 75
4.3.2 Proof . . . 76
4.4 Untrusted network . . . 84
4.5 Considerations and outlook . . . 87

Acknowledgements **91**

Bibliography **92**

Introduction

Quantum cryptography is the science of exploiting the quantum mechanical properties of a quantum system to successfully perform a cryptographic task, such as the secure exchange of information. Before the 1980s quantum cryptography was just a theoretical device with no real application, however this changed when Charles H. Bennett and Gilles Brassard presented the first quantum cryptographic protocol in 1984 ([6]). In their seminal work they applied the idea of "conjugate coding", developed by Stephen Wiesner in the early 1970s: he showed how to encode information in "conjugate observables", of which either one could be received and decoded, but not both ([54]). After this discovery, the field became much more prolific, with many protocols being proposed over the years, and more alive, with the development of the field and its applications to other branches of quantum information theory, computer science and mathematics. This almost explosive interest in quantum cryptography was justified by the advantages that it promised to offer. Up until then in fact the best thing one could do to exchange information in a secure way, was to encrypt it in such a hard cipher, that only someone with the key could decode in a reasonable time.
This is no hard proof of the security of a protocol, and nothing guarantees that one day someone will not find a machine which can efficiently crack encoding schemes that were considered secure, just like *Enigma cipher machine* allowed Poland to decode German communications well before the Second World War. Unlike the classical case, quantum cryptography is built on the solidity of the laws of physics, which are, to the best of our knowledge, unbreakable.
When the need to communicate in a private way arose, namely with the advent of internet for most people, classical cryptography had already been studied for decades (if not centuries), and secure protocols such as the *RSA cryptosystem* ([38]), still used today, were known. On the other hand, the first contributions to the field of quantum cryptography were not even ten years old, and now after more than 40 years, while we keep studying this incredible branch of physics, the field of Quantum Networks developed as well, since it felt like the natural next step in the context of quantum communications.

In this work we propose the first quantum cryptographic protocol native to

quantum networks which does not use input randomness. While Quantum Key Distribution (QKD) and other Quantum Conference Key Agreement (QCKA) protocols already exist, all of them need to use a certain amount of input randomness to produce any amount of output randomness, while our protocol is free from this constraint. We build on recent discoveries in the field of Device-Independent quantum cryptography, in particular applied to the triangle network. Indeed, the discovery of the *RGB4* distribution ([41]) allowed to conjecture that this feat could be possible, having found a self-testing distribution without the need for the parties to randomly choose which measurement to perform each round. Starting from this result, researchers looked into the minimal conditions needed to perform quantum communications in a network, leading to another important discovery, which is that the output randomness of the RGB4 distribution is robust to topological changes in the network. This implies that to perform cryptographic tasks in a network, it is only necessary to trust a small region around oneself and small regions around the other parties involved.
The proof of security of our protocol is based on a reworked version of the *Leftover Hash Lemma* (LHL, [50]), a known and general results of information theory and cryptology, on the relations between von Neumann and one-shot entropies ([34], [55]), and on some recent results on the accumulation of entropy over repeated iterations ([15], [49]). With these we created a cryptographic protocol which can be applied to some different scenarios: M parties who want to create a key inside a network of $N > M$ trusted parties (QKD if $M = 2$, QCKA if $M > 2$) secure against external eavesdroppers, and M parties who want to share a key in a mostly untrusted network (as before QKD if $M = 2$ and QCKA if $M > 2$).

This thesis is organized in four chapters. In Chapter 1 are presented some fundamental concepts and ideas of Quantum Information Theory, from qubits and measurements, to entanglement and nonlocality, and lastly entropy. This chapter is dedicated to explaining the basis over which we will build the next two chapters, with a particular attention to the different kinds of entropies needed and their relations. Chapter 2 is a review of quantum cryptography. Here I will describe some central concepts, such as quantum key distribution, device-independence and classical post-processing, which allow to perform cryptographic tasks. This (simbolically) covers half of the foundational blocks needed to understand the protocol. I will also showcase and describe some famous protocols and quantum attacks to give practical examples. In Chapter 3 is the other half central concepts, in particular I will present quantum networks and their characteristics, along with an in-depth treatment of the RGB4 distribution. Lastly, in Chapter 4 I will present the original part of this project, based on the paper in preparation [2]. The chapter is organized as such: first are explicited the steps for the two-party protocol in a trusted network, then it is adapted to a QCKA protocol, where is also presented

an example for $N = 4$ parties, then is the security proof and finally is the case of a network of unknown topology. The chapter is concluded with some remarks on open questions and possible outlooks.

Chapter 1

Quantum Information Theory

In this first chapter I present the concepts of Quantum Information Theory most important to the project, following mainly [30]. I will start with the notion of qubit to then introduce density operators and distances, then I will present the concept of nonlocality and finally entropy.

1.1 Qubits

As bits are the basis of classical computation and information, quantum bits, or *qubits* for short, are the building blocks of quantum computation and information. Focusing on the mathematical treatment of a qubit, they are interpretable as the elements inhabiting an Hilbert space $\mathcal{H}_1$ spanned by the two vectors $\{|0\rangle, |1\rangle\}$, in Dirac notation, which form an orthonormal basis of states of a system. Differently from classical bits a qubit can be in the *superposition*, mathematically a linear combination, of the two states:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{1.1}$$

meaning that when measuring the state of the qubit we have, consequence of the Born rule, a probability $|\alpha|^2$ of measuring the state $|0\rangle$ and a probability $|\beta|^2$ of measuring the state $|1\rangle$, with the constraints $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. This representation of qubits is not unique however, and another very common representation makes use of the *Bloch sphere*, parametrizing the state using angles and picturing a state as a point on a sphere[1].

[1]The point belongs on the surface of the sphere if the state is normalized and pure, subnormalized states ($|\langle\psi|\psi\rangle|^2 < 1$) and mixed states ($\alpha, \beta \neq 0$) make up the bulk of the sphere.

The representation of a qubit in the Bloch sphere is thus

$$|\psi\rangle = e^{i\gamma}\left(\cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\theta}{2}\right)|1\rangle\right) \tag{1.2}$$

where $\gamma, \theta, \varphi \in \mathbb{R}$ and the term in front can be ignored as it encodes a global effect which is not measurable, so effectively the parametrization of (1.1) only requires two variables, $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ and allows for a one-to-one mapping with the 2-sphere

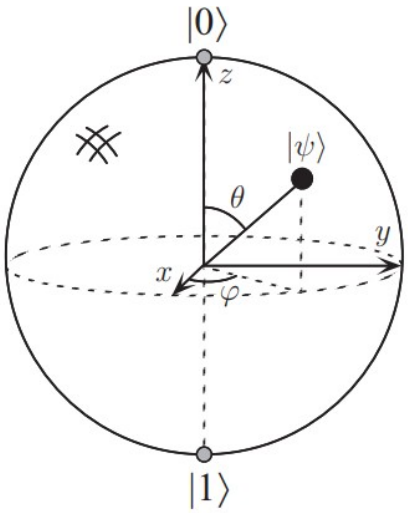

Figure 1.1: Bloch sphere representation of a qubit ([30], p.15)

Another common representation for qubits uses 2-dimensional vectors

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \quad |1\rangle = \begin{pmatrix}0\\1\end{pmatrix} \tag{1.3}$$

which are the eigenstates of the Pauli operator σ_z. The three operators are

$$\sigma_x = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, \quad \sigma_y = \begin{bmatrix}0 & -i\\i & 0\end{bmatrix}, \quad \sigma_z = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix} \tag{1.4}$$

and the eigenstates of σ_x and σ_y are respectively

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}} \quad \text{and} \quad |\pm i\rangle = \frac{|0\rangle \pm i\,|1\rangle}{\sqrt{2}}, \tag{1.5}$$

which are the states representing the points of intersection between the axis and the surface of the Bloch sphere, just as $|0\rangle$ and $|1\rangle$ are the intersections of the z-axis with the surface.

1.1.1 Multiple Qubits

When the system is made up of more than just one qubit it's Hilbert space becomes $\mathcal{H}_N = \mathcal{H}_1^{\otimes N}$. The most general state for a system of two qubits is

$$|\psi\rangle = \alpha_{00}\,|00\rangle + \alpha_{01}\,|01\rangle + \alpha_{10}\,|10\rangle + \alpha_{11}\,|11\rangle \tag{1.6}$$

where the constraints are similar to the single qubit case:

$$\sum_{i\in\{0,1\}^2} |\alpha_i|^2 = 1\ ,\quad \alpha_i \in \mathbb{C} \tag{1.7}$$

with $\{0,1\}^n$ as a short-hand notation for "all the strings of n bits".

1.1.2 Entanglement and Bell States

With the introduction of multiple qubits we can introduce a pivotal concept of quantum mechanics: *entanglement*. Formally it is a resource characteristic of quantum mechanics, however this work is not meant to explore in depth what entanglement is, and limiting myself to a mathematical formulation it can be qualitatively expressed as the impossibility of separating a state into a product of (product) states. This can perhaps be seen easily with an example: let us take the 2-qubit state $|\phi^+\rangle = (|00\rangle + |11\rangle)/2$, there is no way to rewrite it as a product of product states, i.e. only $|0\rangle$ and $|1\rangle$, or more formally

$$\left|\phi^+\right\rangle \neq \sum_j \lambda_j \left|i_j\right\rangle_a \otimes \left|i_j\right\rangle_b \tag{1.8}$$

where $\left|i_j\right\rangle_a, \left|i_j\right\rangle_b$ are basis vectors of the spaces $\mathcal{H}_a$, and $\mathcal{H}_b$ and λ_j are the complex coefficients related to the probabilities of measuring the system in one or the other state. On the contrary the state

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) \tag{1.9}$$

is not entangled, i.e. it is *separable*, because it can be expressed as

$$\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)_A \otimes \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)_B \tag{1.10}$$

The state $|\phi^+\rangle$ is one of four special vectors which make up what is known as the *Bell basis* (1.11), a complete and orthonormal basis of (maximally) entangled vectors spanning the Hilbert space of two qubits

$$\left|\phi^{\pm}\right\rangle = \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}\ , \quad \left|\psi^{\pm}\right\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}} \tag{1.11}$$

which are very important in many protocols, such as superdense coding, and in general in quantum information theory.

1.1.3 Density Operator

State vectors are not the only way to represent the Quantum Mechanics framework, and a different formalism is that of density operators, much more helpful in a variety of cases. A density operator is defined as

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{1.12}$$

where $\{p_i, |\psi_i\rangle\}$ is an *ensemble* of normalized pure states with their associated probabilities, which as always are subject to the constraints $\sum_i p_i = 1,\ p_i \geq 0$. It is also known as a density *matrix* since the expression $|\cdot\rangle\langle\cdot|$ represents the outer product between the two vectors.
In the context of multiqubits systems we can define the reduced density matrix as the density matrix over a space which is not the space of the full system, in fact if $\rho_{AB} \in \mathcal{L}(\mathcal{H}_{AB})$, the reduced density matrix over the space $\mathcal{L}(\mathcal{H}_A)$ is

$$\rho_A = \sum_j p_i p_j \left|\psi_{ij}\right\rangle \left\langle\psi_{ij}\right| = p_i \sum_j \left|\psi_{ij}\right\rangle \left\langle\psi_{ij}\right| \tag{1.13}$$

with $|\psi_{ij}\rangle \in \mathcal{H}_{AB}$.

The density operator is, as the name implies, a linear operator and thus it exists in the space of linear operators acting on the Hilbert space$\mathcal{L}(\mathcal{H})$, however for a density matrix to represent a state there are some properties which must be respected, the proofs of which can be found in [30]:

1. ρ is a positive operator, i.e. all its eigenvalues are real and positive semidefinite.
2. A density operator represents a quantum state $\iff$ Tr[ρ]=1
3. Tr[ρ^2] ≤ 1, saturated if and only if ρ represents a pure state.
4. Density operators are hermitian $\Rightarrow \rho^\dagger = \rho$.

Using these we can restrict the space of the density matrices to

$$S(\mathcal{H}) = \{\rho \in \mathcal{L}(\mathcal{H}) | \rho \geq 0,\ \mathrm{Tr}[\rho] = 1\} \tag{1.14}$$

The density matrix can be expressed in the Bloch formalism as well

$$\rho = \frac{\mathbb{1} + \vec{r} \cdot \vec{\sigma}}{2} \tag{1.15}$$

where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is a three-dimensional vector whose elements are the Pauli matrices and $\vec{r}$ is the Bloch vector relative to the state.

1.1.4 Fidelity and Trace Distance

To evaluate how different two quantum states are one can use two main *distances*, the trace distance $D(\rho, \tau)$ or the fidelity, which are related one another. The trace distance is defined as

$$\frac{1}{2}||\rho - \tau||_1 = \frac{1}{2}\mathrm{Tr}\big[|\rho - \tau|\big] \tag{1.16}$$

which has a nice geometrical interpretation with the Bloch sphere. Take the Bloch representation of the density matrix (1.15) assigning $\vec{r}$ to ρ and $\vec{s}$ to τ, then the trace distance becomes

$$D(\rho, \tau) = \frac{1}{4}\mathrm{Tr}|(\vec{r} - \vec{s}) \cdot \vec{\sigma}| = \frac{|\vec{r} - \vec{s}|}{2} \tag{1.17}$$

where the second equality comes from the fact that the Pauli matrices all have eigenvalues ± 1 so the eigenvalues of $(\vec{r} - \vec{s}) \cdot \vec{\sigma}$ are $\pm(\vec{r} - \vec{s})$. In this way we can see that geometrically the trace distance between two states is literally the distance in the Bloch sphere between the two vectors representing the states.

The fidelity is a measure of how much two states overlap and is defined as

$$F(\rho, \sigma) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right] \tag{1.18}$$

For practical reasons this is not the form of fidelity we used, so following is the Uhlmann theorem, which gives a more general formulation

Theorem 1.1.1 (**Uhlmann's theorem**). *Suppose ρ and σ are states of a quantum system Q. Introduce a second quantum system R which is a copy of Q. Then*

$$F(\rho, \sigma) = \max_{|\Psi\rangle, |\Phi\rangle} |\langle \Psi | \Phi \rangle|, \tag{1.19}$$

where the maximization is over all purifications $|\Psi\rangle$ of ρ and $|\Phi\rangle$ of σ into RQ.

The proof of the theorem is in Chapter 9.2.2 of [30]. A purification of a state is simply a pure state in an Hilbert space that contains the Hilbert space of the state to be purified, and such that the reduced density matrix over the "extra" space gives the state to be purified. Formally this is expressed as

Definition 1.1.1 (**Purification of a state**)**.** Given $\rho = |\psi\rangle\langle\psi|$ with $|\psi\rangle \in \mathcal{H}_A$ and given $\bar{\rho} = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle \in \mathcal{H}_{AB} \supseteq \mathcal{H}_A$, $|\Psi\rangle$ is a purification of $|\psi\rangle$

$$\iff \rho = \mathrm{Tr}_B[\bar{\rho}] \tag{1.20}$$

The inner product $\langle\psi|\phi\rangle$ is known as the overlap between two states and the fidelity is

$$F^2(\rho, \sigma) = |\langle\psi|\phi\rangle|^2 \tag{1.21}$$

since the overlap is a complex number.

1.2 Measurements

To extract some value or quantity from a state one has to perform a measurement, and in quantum mechanics measurements are described by operators, usually called $\{M_m\}$, acting on the system which is being measured. Formally

Definition 1.2.1 (**Measurements**)**.** Quantum measurements are described by collections of operators $\{M_m\}$, where $M_m : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ and $\mathcal{B}(\mathcal{H})$ is the set of linear bounded operators acting on the Hilbert space of the system. The index m is related to the possible outcomes of the measurement.

In general there are multiple possible outcomes of a measurement, and from this follows that we have a probability associated to each outcome m. For example the probability of getting a specific value $\overline{m}$ measuring a state $|\psi\rangle$ with a measurement $\{M_m\}$ is

$$p(\overline{m}) = \langle\psi| M_{\overline{m}}^\dagger M_{\overline{m}} |\psi\rangle = |M_{\overline{m}} |\psi\rangle|^2 \tag{1.22}$$

and after a measurement the state will be

$$|\psi'\rangle = \frac{M_m |\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m |\psi\rangle}} \tag{1.23}$$

Since we have a relation between probabilities and measurement operators, there must be some kind of constraint related to the fact that the sum of all

probabilities must amount to 1. This translates into the fact that a set of operators to qualify as a measurement must be a *complete* set of operators, meaning

$$\sum_m M_m^\dagger M_m = \mathbb{1} \tag{1.24}$$

In terms of density operators we have

$$p(m) = \mathrm{Tr}[M_m^\dagger M_m \rho], \ , \quad \rho' = \frac{M_m \rho M_m^\dagger}{\sqrt{\mathrm{Tr}[M_m \rho M_m^\dagger]}} \tag{1.25}$$

An important aspect of measuring is, other than the trivial one of having some result to check, the possibility of distinguishing quantum states. One can however **reliably** distinguish only quantum states which are orthogonal to one another. A proof by contradiction is the following:

Assume there are two non-orthogonal ($\langle\psi_1|\psi_2\rangle > 0$) states $|\psi_1\rangle$ and $|\psi_2\rangle$, and suppose there exists a measurement $\{M_j\}$ which can consistently distinguish them. If we were to measure an unknown state between $|\psi_1\rangle$ and $|\psi_2\rangle$ this means that there is some deterministic function f such that, once an outcome j is gathered, $f(j)$ will unequivocally point to the correct initial state. Such observations can be written as

$$\langle\psi_1|E_1|\psi_1\rangle = 1 \ , \quad \langle\psi_2|E_2|\psi_2\rangle = 1 \tag{1.26}$$

with $E_i = \sum_{j:f(j)=i} M_j^\dagger M_j$.
Since these are operators they must still respect $\sum_i E_i = \mathbb{1}$, thus $\sum_i \langle\psi_1|E_i|\psi_1\rangle = 1$ as well, and since $\langle\psi_1|E_1|\psi_1\rangle = 1$ this implies $\langle\psi_1|E_2|\psi_1\rangle = 0$. At this point we can notice that the operators E_i are both positive and hermitian so if $\langle\psi_1|E_2|\psi_1\rangle = 0$ then $\sqrt{E_2}\,|\psi_1\rangle = 0$ too. We can now decompose $|\psi_2\rangle$ into a component parallel to $|\psi_1\rangle$, since they're not orthogonal, and a component perpendicular to it, so $|\psi_2\rangle = \alpha\,|\psi_1\rangle + \beta\,|\varphi\rangle$ with of course $\alpha^2 + \beta^2 = 1$ and $|\beta| < 1$. The last passage is to uncover the contradiction: $\sqrt{E_2}\,|\psi_2\rangle = \beta\sqrt{E_2}\,|\varphi\rangle$ thus $\langle\psi_2|E_2|\psi_2\rangle = |\beta|^2\langle\varphi|E_2|\varphi\rangle \leq |\beta|^2 < 1$, in contrast to (1.26).

Two special kinds of measurements, which are often used in many quantum protocols are the *projective* measurements and the *positive operator-valued* measure, or POVM.

Definition 1.2.2 (**Projective measurements**)**.** A projective measurement is described by an observable, i.e. an Hermitian operator on the state space of the system being observed, and it has the spectral decomposition

$$M = \sum_m m P_m \tag{1.27}$$

where P_m is a projector onto the eigenspace of M with eigenvalue m. The projectors are orthogonal to one another, make up a complete set of operators and are idempotent, i.e. $P_m P_n = \delta_{mn} P_m$ and $\sum_m P_m = \mathbb{1}$.

Definition 1.2.3 (**Positive Operator-Valued Measure**)**.** A POVM is a collection of operators $\{E_i\}$ where $E_i = M_i^\dagger M_i$ satisfying

$$\{E_i\} \succeq 0 \ , \ \sum_i E_i = \mathbb{1} \tag{1.28}$$

where $\succeq$ means element-wise semi-definite positiveness.

In Section 2.2.8 of [30] is proven that projective measurement when enhanced with unitary operations are equivalent to general measurements.

1.3 The EPR paradox and Nonlocality

Entanglement is strictly related[2] to another central concept of quantum mechanics, which is *nonlocality*. To understand what nonlocality is we need to take a step back to 1935, when Albert Einstein, Boris Podolski and Nathan Rosen published a paper ([16]) where they argued that quantum mechanics cannot be a complete theory. In fact, after contributing to formulating quantum mechanics, Einstein was so flabbergasted by developments such as entanglement and nonlocality, which he famously defined as "spooky action at a distance", that he assumed quantum mechanics must have been an incomplete theory. Up until that point a complete theory needed to have two characteristics: it needed to be *local*, meaning that events cannot be correlated if the interval between them is space-type (i.e. the signal would need to travel faster than light), and *realistic*, meaning that the object of observation exists disregarding of the measurement happening. Quantum mechanics violated (allegedly) both of these principles, and we can see it with

[2]Entanglement is largely considered to be a nonlocal effect, but it is possible to have nonlocality without entanglement [8].

a simple thought experiment: A and B share two entangled particles and then travel away from each other, potentially light-years away, once they reach their destinations A measures the spin of their particle and instantly knows what would be the outcome of B's measurement, even without them measuring their particle. It is easy to see how Einstein, Podolski and Rosen arrived at the conclusion that, since nothing travels faster than light, the outcomes of quantum measurements must be elements of reality, existing even when not measured. The fact that it's not possible to predict these outcomes must be due to a set of hidden variables, unknown to us and such that, if revealed, would allow to make predictions and make quantum mechanics a complete theory. Unfortunately for them there are some fallacies in this reasoning, specifically the information did not travel faster than light because B does not know the outcome of his (hypotetical) measurement until A tells them, and of course that kind of communication is subject to causal constraints. The natural follow-up question would be:
But what if B actually does the measurement instead of waiting for A to tell them? Could we say that causality was violated in that case?
The answer is again no, in fact for A and B to share two entangled particles the event of entangling them must be in the shared region of their past light cones (Fig. 1.2), so the information is "stored" in the particles in a moment where the two parties were in a time-type interval (i.e. a distance which allows causality between events).

1.3.1 Bell Inequalities

To have a first formal disproof of the EPR paradox the community had to wait almost 30 years, when John Stewart Bell published his pivotal paper proving that no local hidden variable (LHV) model is compatible with quantum mechanics ([5]). In the paper, Bell states that if there is this LHV model then it's possible to find some constraints which take the name of Bell-type inequalities, there are (infinitely) many different Bell-like inequalities, all equivalent, and proving that one is violated is enough to prove that quantum mechanics is a complete theory. Bell's formulation was not easy to reproduce experimentally and almost 10 years later there was the first *Bell test*, executed by John Francis Clauser and Stuart Freedman, and after that many Bell tests were performed optimizing the setups, closing loopholes and verifying violations of different inequalities.

1.3.2 The Bell theorem and the CHSH inequality

Bell's treatment in [5] is overly complicated and definitely not the most common one, so here is a more digestible version of Bell's theorem and the CHSH (Clauser-Horne-Shimony-Holt) inequality.

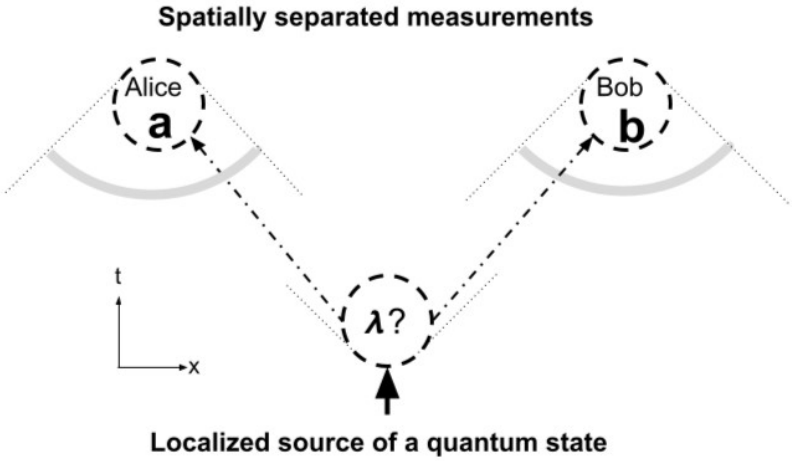

Figure 1.2: On the horizontal axis is space and on the vertical axis is time. The event λ which entangled the particles that now A and B have must be in a past common to both. (Credits: Johnjbarton - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=138630695)

Theorem 1.3.1 (**Bell's Theorem**). *There exist* $\vec{a}_1, \vec{a}_2, \vec{b}_1, \vec{b}_2 \in \mathbb{R}^3$ *with unitary norm such that no joint probability distribution for the outcomes of the measurement of* $A_i = \vec{a}_i \cdot \vec{\sigma}_A$ *and* $B_i = \vec{b}_i \cdot \vec{\sigma}_B$ *is compatible with the predictions of the quantum mechanics state* $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$.

Before proving this it's necessary to prove

Lem 1.3.1. *Taken any four random variables* $x_1, x_2, y_1, y_2 \in \{+1, -1\}$ *we have* $x_1(y_1 + y_2) + x_2(y_1 - y_2) \leq 2$

Proof.

$$\begin{aligned} &x_1(y_1 + y_2) + x_2(y_1 - y_2) \leq |x_1(y_1 + y_2) + x_2(y_1 - y_2)| \leq \\ &\leq |x_1| \cdot |y_1 + y_2| + |x_2| \cdot |y_1 - y_2| = 2 \\ &\Rightarrow E[x_1(y_1 + y_2) + x_2(y_1 - y_2)] \leq 2 \end{aligned} \tag{1.29}$$

Where $E[\cdot]$ is the expectation value. □

This is what is known as the Bell inequality and it's not intrinsically related to quantum mechanics, but its violation is what guarantees the presence of nonlocality (i.e. the violation of the CHSH inequality is a nonlocality witness). Going back

to Theorem 1.3.1, assume x_1, x_2, y_1, y_2 to be the random variables associated to the four operators A_1, A_2, B_1, B_2, the first two are a measurement that A performs and the alst two are a measurement that B performs.

Proof. Using the linearity of the expectation value we can rewrite the last line of (1.29) as

$$E[x_1y_1] + E[x_1y_2] + E[x_2y_1] - E[x_2y_2] \tag{1.30}$$

and since these are all expectation values of compatible observables, meaning that the measurement of these quantities can be done simultaneously since the operators are on different spaces (A_i belong to A and B_i belong to B), they can be evaluated on quantum states.

$$\begin{aligned} \Rightarrow &\left\langle \psi^- \right| A_1B_1 \left| \psi^- \right\rangle + \left\langle \psi^- \right| A_1B_2 \left| \psi^- \right\rangle + \left\langle \psi^- \right| A_2B_1 \left| \psi^- \right\rangle - \left\langle \psi^- \right| A_2B_2 \left| \psi^- \right\rangle = \\ &= \left\langle \psi^- \right| A_1(B_1 + B_2) \left| \psi^- \right\rangle + \left\langle \psi^- \right| A_2(B_1 - B_2) \left| \psi^- \right\rangle = \\ &= \left\langle \psi^- \right| \vec{a}_1 \cdot \vec{\sigma}_A \otimes (\vec{b_1} + \vec{b_2}) \cdot \vec{\sigma}_B \left| \psi^- \right\rangle + \left\langle \psi^- \right| \vec{a}_2 \cdot \vec{\sigma}_A \otimes (\vec{b_1} - \vec{b_2}) \cdot \vec{\sigma}_B \left| \psi^- \right\rangle \end{aligned} \tag{1.31}$$

Next choose $\vec{a}_1 = \hat{e}_1, \vec{a}_2 = \hat{e}_2, \vec{b}_1 = -\frac{\hat{e}_1+\hat{e}_2}{\sqrt{2}}, \vec{b}_2 = \frac{\hat{e}_2-\hat{e}_1}{\sqrt{2}}$, so that $\vec{b}_1 + \vec{b}_2 = -\sqrt{2}\hat{e}_1$ and $\vec{b}_1 - \vec{b}_2 = -\sqrt{2}\hat{e}_2$ where $\hat{e}_1$ and $\hat{e}_2$ are two generic axis along which the measurement is made. It's easy to see that $\vec{a}_1 \perp \vec{a}_2$ and $\vec{b}_1 \perp \vec{b}_2$, geometrically we have that these four vectors are coplanar, which maximizes the violation of the inequality. This yields the violation of the CHSH inequality

$$-\sqrt{2} \left\langle \psi^- \right| \hat{e}_1 \cdot \vec{\sigma}_A \otimes \hat{e}_1 \cdot \vec{\sigma}_B \left| \psi^- \right\rangle - \sqrt{2} \left\langle \psi^- \right| \hat{e}_2 \cdot \vec{\sigma}_A \otimes \hat{e}_2 \cdot \vec{\sigma}_B \left| \psi^- \right\rangle = 2\sqrt{2} \tag{1.32}$$

where we used that $|\psi^-\rangle$ has the two bits anticorrelated and is antisymmetric, thus the expectation value of a generic unitary-norm vector on the two qubits will be -1. □

To recap, this violation implies that no joint probability distribution for the outcomes of these measurements is possible, meaning that no LHV model is compatible with quantum mechanics.

1.3.3 Polytopes

A compact way to visualize whether a probability distribution is local or not, is through the use of *polytopes*, geometric object which generalize polyhedra to higher dimensions, commonly used in the context of non-locality. Typically a Bell polytope is divided in regions, from the most external to the most internal:

- **No-signalling (NS)** region: all the distributions inside the NS region respect causality, meaning that the outcome of a measurement of a party does not influence the outcome of the measurement of another party unless they share a common past where some correlation happened.
- **Quantum** region: the distributions in this region can be reproduced doing measurements on bipartite quantum systems.
- **Local** region: the distributions inside this region can be reproduced using only shared randomness as a resource.

In Fig. 1.3 is an example of a generic polytope. As we can see only the local and NS regions are polytopes in and of itself, while the quantum region is not since it has some curved facets, or more formally it has an infinite number of extremal points.

To understand better we can consider the standard Bell scenario, indicated as (2, 2, 2), since there are two parties with two possible measurements each which give binary outcomes. Starting from the characterization of $\mathcal{P}$ we have the no-signaling condition, which can be translated mathematically as

$$\begin{aligned} \sum_b p(a,b|x,y) &= p(a|x) \quad \forall a,b,x,y \\ \sum_a p(a,b|x,y) &= p(b|y) \quad \forall a,b,x,y \end{aligned} \tag{1.33}$$

which individuates the vertices of the polytope. This set of equations is respected by 24 total probability distributions, of which 16 represents the $\mathcal{L}$ vertices and 8 the non-local vertices. I will not go into detail into the derivation of all the vertices, but it may be useful to see an example: in the literature the non-local vertices are called PR boxes (Popescu-Rohrlich) and represents correlation which are no-signaling but unattainable with quantistic correlations, in fact these particular

[3]The space containing no-signaling distributions that violate the Tsirelson bound (CHSH$\leq 2\sqrt{2}$).

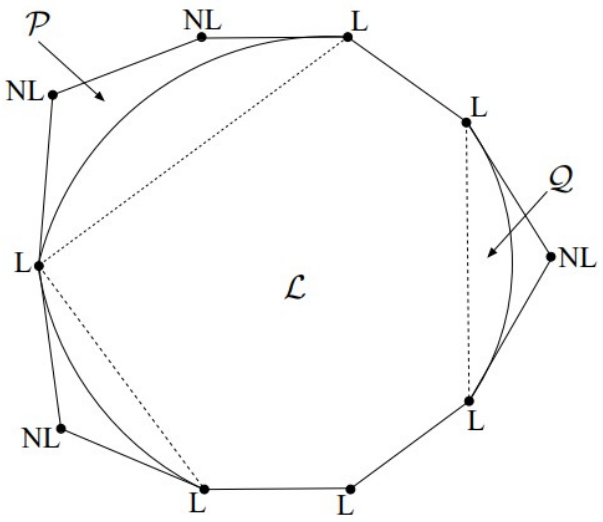

Figure 1.3: $\mathcal{P}$ is the No-Signaling polytope, $\mathcal{Q}$ is the Quantum region and $\mathcal{L}$ is the Local region. Bell inequalities are the hyperplanes (dotted lines) separating the local and quantum region, while the hypersurfaces separating the quantum and the no-signaling region are the Tsirelson bounds, super-quantum correlations[3]. ([4], p.3)

distributions allow to win a CHSH-like game[4] with probability 1. Classically the best thing the players can do is to always output 0 because this strategy gives a 75% probability of winning, while if we allow quantum correlations the upper bound in the probability will not be related to the Bell inequality, but to the Tsirelson bound, which allows for a maximum winning probability of ≈ 0.853 (the proof is in Chapter 4.3.3 of [33]). Allowing super-quantum correlations it's possible to achieve 100% probability of winning without signaling, and a distribution of this kind (also a vertex of the $\mathcal{P}$ polytope) is

$$p(a, b|x, y) = \frac{1}{2}\delta_{a\oplus b, x\wedge y} \tag{1.34}$$

This is indeed no-signaling, in fact we can show that the probability distribution is independent of y for the party A (it works in the same way for B). The distribution can be represented in matrix form as

$$\begin{bmatrix} p(a=0,b=0,|x,y) & p(a=1,b=0,|x,y) \\ p(a=0,b=1,|x,y) & p(a=1,b=1,|x,y) \end{bmatrix}$$

[4]A CHSH game is a "challenge" that two parties can perform where they have to output $a = \neg\, b$ if $x = y = 1$, and $a = b$ in all the other cases (i.e. they win if $a \oplus b = x \wedge y$)

- if $x = 0$ then $x \wedge y = 0 \; \forall y$ thus

$$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

This summed over a gives $p(b|x = 0, y)$ and since the matrix does not change for $y = 0$ and $y = 1$ the result is the same, proving that the marginals are independent of the specific value y takes.

- if $x = 1$ and $y = 1$ we have

$$\begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$$

which summed over a gives $p(b, |x = 1, y = 1)$

$$\begin{aligned} p(a = 0, b = 0|x = 1, y = 1) + p(a = 1, b = 0|x = 1, y = 1) &= 0 + \frac{1}{2} \\ p(a = 0, b = 1|x = 1, y = 1) + p(a = 1, b = 1|x = 1, y = 1) &= \frac{1}{2} + 0 \\ \Rightarrow p(b, |x = 1, y = 1) &= \left(\frac{1}{2}, \frac{1}{2}\right) \end{aligned} \tag{1.35}$$

- For $x = 1$ and $y = 0$ we have

$$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

that summed over a gives $p(b|x = 1, y = 0)$

$$\begin{aligned} p(a = 0, b = 0|x = 1, y = 0) + p(a = 1, b = 0|x = 1, y = 0) &= \frac{1}{2} + 0 \\ p(a = 0, b = 1|x = 1, y = 0) + p(a = 1, b = 1|x = 1, y = 0) &= 0 + \frac{1}{2} \\ \Rightarrow p(b, |x = 1, y = 0) &= \left(\frac{1}{2}, \frac{1}{2}\right) \end{aligned} \tag{1.36}$$

Proving that the marginals are the same in this case too.

The non-local vertices are all reducible to the distribution in (1.34) up to reversible local operations (a relabeling of the inputs and outputs).
The quantum set is not a polytope[5], which means that it's not possible to characterize it through its vertices since there's an infinite number of them.

The local set is the convex hull of the polytope identified by the intersection of all the Bell inequalities. These are in fact hyperplanes, and we can see it by noticing that a generic probability distribution of the kind $p(a, b|x, y)$ is a 16-dimensional vector (2 parties with 2 measurements with binary outcomes), using the 4 normalization conditions

$$\sum_{a,b} p(a, b|x, y) = 1 \ \ \forall x, y \tag{1.37}$$

and the 4 no-signaling conditions of (1.33) we get 8-dimensional vectors. Since Bell inequalities take the form

$$\sum_{a,b,x,y} c_{a,b,x,y} p(a, b|x, y) \leq \mathcal{S} \tag{1.38}$$

this removes one degree of freedom, thus in the space of the polytopes these are $(d-1)$-dimensional hyperplanes confining the local set of correlations; they are all the same inequality up to relabeling, in fact there is only one tight Bell inequality in the CHSH scenario, which is of course the CHSH inequality, as stated in Chapter 2.5.1 of [39]. The vertices of the local polytope are, as mentioned before, also part of the no-signaling polytope, this is because they are all (and exclusively, since any other local distributions can be written as a convex combination of these) the 16 deterministic functions that associate a fixed value to the outputs of the parties independently of all the other variables.

An important feature of Bell polytopes in the standard Bell scenario is that they are convex, meaning that taken $t \in [0, 1]$ and any two probability distributions $p_1, p_2 \in \mathcal{L}$ the linear combination $tp_1 + (1 - t)p_2$ is in $\mathcal{L}$, and this is valid for $\mathcal{P}$ as well as for $\mathcal{Q}$[6], a feature that will not be present in quantum networks, as I will show later.

[5]The proof that the quantum set is not a polytope (but an elliptope) is overly complex and long, but it can be found in [26], with an explicit analytical derivation of the maximization of the CHSH done in the appendix of [19]

[6]The set of quantum correlations is in general convex, but there are some exceptions, in our (2,2,2) case it's convex since it has one input per party. For details refer to [12] (there is a slight mismatch in notation, in this work the number of inputs is global, while in the cited paper the number of inputs is for each party).

1.4 Entropy

Entropy is a central concept in both classical and quantum information theory, and as such it has raised much interest in the scientific community as a whole. A general (and generalistic) definition of entropy, which is often popularized, is that entropy is a measure of disorder or of chaos in a system. I think that if we had to limit ourselves to general considerations entropy has a much more fascinating interpretation of a measure of the lack of information that we have about a system, following the ideas of E. T. Jaynes. The concepts of "order" and "disorder" are tightly related to our experience as humans, signifying respectively "something that is easy to predict/describe" and "something that is too complex to find a pattern into", but maybe they're not the fundamental concepts to focus on, and I give this role to the "idea" of information. An enlightening discussion of this interpretation is given in [13] which I will summarize in the rest of this introductory paragraph.
When talking about thermodynamics and entropy, the original framework by Clausius does not explicitly connect with information theory at all, and in fact the first "visible" connection came with the definition of entropy by Shannon $E_{SH} \propto -\sum_i p_i \log_2 p_i$, which is formally equal to Gibbs' entropy $E_G = -k_B \sum_i p_i \ln p_i$. Mathematical similarity however, never was enough to guarantee a deeper link between physical concepts, and a stronger connection between entropy and information was given by Jaynes ([23], [24]), where he showed that the usual thermodynamic results emerge naturally with an information theoretic approach. For the sake of completeness it is only fair to recognize that before Jaynes came Szilard in 1929 with the famous Maxwell's demon thought experiment, which drove many scientists to explore the similarities between thermodynamics and information dynamics. A particularly beautiful result was achieved by Rolf Landauer when he formulated the *Landauer's principle*, stating that erasing (1 bit of) information (i.e. resetting a system to it initial state) is an operation which inherently increases the entropy (by at least $k_b \ln 2$). The fascinating aspect of all this is that we are left with two choices: either the abstract concept of information is not as abstract and subjective as we are used to think, being subject to thermodynamic laws, or, if we believe that information and thermodynamics are independent and unrelated, then it can be used to violate the second law of thermodynamics.
Then the question that comes up naturally is: was the second law about information all along?
In most of physics books and courses the formulations of thermodynamic and entropy are made without ever mentioning information dynamics, but only from microscopic considerations, such as Hamiltonian and Lagrangian mechanics in classical physics and unitary quantum mechanics in modern physics. Disregarding of

how different approaches use different assumptions and approximations, one thing that is common to all attempts of deriving thermodynamics without using information theory, is that there is always a passage which can be interpreted as a loss of information. Some examples are stochastic approximation, random phase approximation and coarse graining, which all introduce "new physics" into the model outside[7] of the energy dynamics, and the bottom line is that to make the connection between microscopic energy behaviors and the second law of thermodynamic we must throw away information, allowing us to infer that in nature information is lost. We can rephrase the second law fo thermodynamics as

"No process can result in a net gain of information."

This does not answer the question "why entropy increases?" (or in this case "why is information lost?"), but shifts the focus on something that is (maybe) more fundamental, and that will hopefully help us get a clearer understanding of the world around us[8].
In this spirit I will present entropies of all shapes and forms trying to always tie them to some kind of information loss/gain process, which I hope will also make the rest of this thesis a more easy read.

1.4.1 Classical Entropies

Classical entropy refers to classical information and was first formulated by Claude Shannon in 1948 [42], the *Shannon entropy* of a random variable X can be defined as the uncertainty we had on X before learning its value, following is a mathematical definition.

Definition 1.4.1. Shannon entropy: The entropy of a discrete random variable X with probability distribution $p_X(x)$ is defined as

$$H(X) = -\sum_x p_X(x) \log(p_X(x)) \tag{1.39}$$

The expression log always refers to $\log_2$, reserving for $\log_e$ the common notation ln, and by reason of $\lim_{\varepsilon \to 0^+} \varepsilon \log \varepsilon = 0$ we take the convention $0 \cdot \log 0 = 0$. This definition is a result of *Shannon's noiseless channel coding theorem*,

[7] It cannot be inside since from Liouville's theorem the conservation of phase-space volume under time evolution implies no loss of information on the system.

[8] There are still many open questions and this interpretation is far from complete, for example the Entropy Fluctuation Theorem (a closed system's entropy has a non-zero probability of spontaneously decreasing) and Loschmidt's paradox (it should be impossible to deduce irreversible processes, such as the increase of entropy, from time-symmetric dynamics, such as most low-level fundamental physical process) seem to violate the second law, indicating that a deeper understanding is necessary.

Theorem 1.4.1 (**Shannon's noiseless channel coding theorem**). *Given* $\{X_i\}$ *identically and independently distributed (IID) sources of classical information with entropy rate* $H(X)$*, if* $R > H(X)$ *there exist a reliable compression scheme of rate* R*.*

The proof of this theorem is in Chapter 12.2.1 of [30], with a detailed explanation of all the quantities in this expression, here I will give a more simplified (and qualitative) explanation to bridge together (1.39) with the interpretation of entropy as lack of information. For now we can take axiomatically (will be proven later) that $H(X) \leq \log |\mathcal{X}|$, where $\mathcal{X}$ is the alphabet (all the possible values) that X can take, and a compression scheme Φ is a map that sends a bitstring of length n into a compressed bitstring of length nR ($\lfloor nR \rfloor$ for non-integer R); in this context the entropy of X are the minimal physical resources necessary and sufficient to reliably store the output from the source. Assume now that $\{X_i\} = X_1, X_2, X_3, \ldots$ are IID random variables taking values 0 with probability p and 1 with probability $1-p$, taken a sequence of values $(x_1, x_2, 1, \ldots, x_n)$ it is reasonable to expect that as $n \to +\infty$ there will be a fraction p of 0s and a fraction $1-p$ of 1s, mathematically

$$p(x_1, \ldots, x_n) = p(x_1) \ldots p(x_n)\Big|_{n\to+\infty} \approx p^{np}(1-p)^{n(1-p)} \tag{1.40}$$

Taking the logarithm on both sides provides us with the definition

$$-\log p(x_1, \ldots, x_n) \approx -np \log p - n(1-p)\log(1-p) \stackrel{def}{=} nH(X) \tag{1.41}$$

which gives rise to another important and common formula:

$$p(x_1, \ldots, x_n) \approx 2^{-nH(X)} \tag{1.42}$$

In this case ($|\mathcal{X}| = 2$) the entropy has a special name and it's called *binary entropy*

$$h(X) = -p_X(x) \log p_X(x) - (1 - p_X(x)) \log(1 - p_X(x)) \tag{1.43}$$

1.4.1.1 Mathematical Properties

A trivial property of the Shannon entropy, given by the definition and the properties of logarithms, is

$$H(X) \geq 0 \tag{1.44}$$

since probabilities are $0 \leq p_X(x) \leq 1$.

Another property is the *concavity*

$$H(\lambda p_1 + (1-\lambda)p_2) \geq \lambda H(p_1) + (1-\lambda)H(p_2) \tag{1.45}$$

but to prove this we need to introduce the relative entropy.

Definition 1.4.2 (**Relative Entropy**)**.** Suppose $\mathcal{X}$ is a finite set and $p(x)$ and $q(x)$ are two probability distributions over the set $\mathcal{X}$. The relative entropy (or the Kullback-Leibler divergence) of $p(x)$ to $q(x)$ is

$$D\big(p(x)||q(x)\big) = \sum_{x\in\mathcal{X}} p(x)\log\left(\frac{p(x)}{q(x)}\right) = -H(x) - \sum_{x\in\mathcal{X}} p(x)\log q(x) \tag{1.46}$$

and can be interpreted as the error we would commit if we approximated a distribution $p(x)$ with $q(x)$ instead.
So, to prove now that $H(x)$ is concave,

Proof. Let $q(x)$ be the uniform distributions over the set $\mathcal{X}$ with $|\mathcal{X}| = d$, then

$$H(x) = -D\big(p(x)||q(x)\big) - \sum_{x\in\mathcal{X}} p(x)\log(1/d) = -D\big(p(x)||q(x)\big) + \log d \tag{1.47}$$

Proving that $H(x)$ is concave reduces to proving that $D\big(p(x)||q(x)\big)$ is convex, given that if $f(x)$ is convex $-f(x)$ is concave and we are only adding a constant to that. From Def. 1.46

$$D\big(p(x)||q(x)\big) = \sum_{x\in\mathcal{X}} p(x)\log\left(\frac{p(x)}{q(x)}\right) \geq \left(\sum_{x\in\mathcal{X}} p(x)\right)\log\frac{\sum_{x\in\mathcal{X}} p(x)}{\sum_{x\in\mathcal{X}} q(x)} \tag{1.48}$$

where the inequality is due to the *log sum inequality*. This implies

$$\begin{aligned}
&D\big(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2\big) = \\
&= \sum_{x\in\mathcal{X}}\left[\lambda p_1(x) + (1-\lambda)p_2(x)\cdot\log\frac{\lambda p_1(x) + (1-\lambda)p_2(x)}{\lambda q_1(x) + (1-\lambda)q_2(x)}\right] \leq \\
&\leq \sum_{x\in\mathcal{X}}\left[\lambda p_1(x)\log\frac{\lambda p_1(x)}{\lambda q_1(x)} + (1-\lambda)p_2(x)\log\frac{(1-\lambda)p_2(x)}{(1-\lambda)q_2(x)}\right] = \\
&= \lambda\sum_{x\in\mathcal{X}} p_1(x)\log\frac{p_1(x)}{q_1(x)} + (1-\lambda)\sum_{x\in\mathcal{X}} p_2(x)\log\frac{p_2(x)}{q_2(x)} = \\
&= \lambda D\big(p_1(x)||q_1(x)\big) + (1-\lambda)D\big(p_2(x)||q_2(x)\big)
\end{aligned} \tag{1.49}$$

□

Associated to this we can also state

$$D\big(p(x)||q(x)\big) \geq 0 \tag{1.50}$$

the proof of which is in Chapter 3.1.5 of [55], and we can now prove

$$H(X) \leq \log |\mathcal{X}| \tag{1.51}$$

Proof.
The proof is derived trivially from (1.50) and (1.47). The inequality is saturated $\iff$ $p(x)$ is the uniform distribution. □

1.4.1.2 Joint and Conditional Entropy

The joint entropy is associated to the information related to the pair of random variables X, Y and is defined as

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y) \tag{1.52}$$

and the Shannon entropy has the property of being *subadditive*, meaning that

$$H(X,Y) \leq H(X) + H(Y) \tag{1.53}$$

Proof. Starting from the relative entropy we have

$$\begin{aligned} D\big(p(x,y)||p(x)p(y)\big) &= \sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)} = \\ &= -H(p(x,y)) - \sum_{x,y} p(x,y) \log p(x) - \sum_{x,y} p(x,y) \log p(y) = \\ &= -H(p(x,y)) - \sum_{x} p(x) \log p(x) - \sum_{y} p(y) \log p(y) = \\ &= -H(p(x,y)) + H(p(x) + H(p(y) = \\ &= -H(X,Y) + H(X) + H(Y) \end{aligned} \tag{1.54}$$

which together with (1.50) completes the proof. □

For later use here is the definition of *mutual information* as well

Definition 1.4.3 (**Mutual Information**)**.**

$$H(X:Y) = H(Y) - H(X,Y) \tag{1.55}$$

but I will not explore this quantity any further, details on this can be found in Chapter 3.1.4 of [55] and in Chapter 11.2.3 of [30].

Using the joint entropy it's possible to formulate the *conditional* entropy, which is an important tool representing the amount of information we'd get on a variable X knowing the value of a variable Y.

Definition 1.4.4 (**Conditional Entropy**)**.** Let X and Y be discrete random variables with joint probability distribution $p(x, y)$, then

$$\begin{aligned} H(X|Y) &= -\sum_y p(y) H(X|Y=y) = -\sum_y p_y \sum_x p(x|y) \log p(x|y) = \\ &= -\sum_{x,y} p(x,y) \log p(x|y) = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)} = \\ &= -\sum_{x,y} p(x,y) \log p(x,y) + \sum_{x,y} p(x,y) \log p(y) = \\ &= H(X,Y) - H(Y) \end{aligned} \tag{1.56}$$

The conditional entropy is positive, and we can show it using the Def. 1.4.3

$$\begin{aligned} H(X|Y) = H(X) - H(X:Y) &= -\sum_x p(x) \log p(X) - \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \\ &= -\sum_{x,y} p(x,y) \left[\log p(x) + \log \frac{p(x,y)}{p(x)p(y)} \right] = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)} = \\ &= D\big(p(x,y)||p(y)\big) \geq 0 \end{aligned} \tag{1.57}$$

An important point is that conditioning cannot increase the entropy, meaning

$$H(X) \geq H(X|Y) \tag{1.58}$$

Proof.

$$H(X|Y) = H(X,Y) - H(Y) \leq H(X) \tag{1.59}$$

using (1.54). The inequality is saturated $\iff$ X and Y are independent. □

We can refine the notion of subadditivity to the *strong subadditivity* of Shannon entropy

$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z) \tag{1.60}$$

Proof.

$$H(X,Y,Z) = -\sum_{x,y,z} p(x,y,z)\log p(x,y,z) =$$
$$= -\sum_{x,y,z} p(x,y,z)\log p(x|y,z)p(y,z) = H(Y,Z) - \sum_{x,y,z} p(x,y,z)\log p(x|y,z) \tag{1.61}$$

$$\begin{aligned}\Rightarrow & H(X,Y,Z) - H(Y,Z) - H(X,Y) + H(Y) = \\ & -\sum_{x,y,z} p(x,y,z)\log p(x|y,z) + \sum_{x,y,z} p(x,y,z)\log p(x|y) = \\ & = \sum_{x,y,z} p(x,y,z)\log\frac{p(x|y)}{p(x|y,z)}\end{aligned} \tag{1.62}$$

Now use $\log x \leq (x-1)/\ln 2$

$$\begin{aligned}\Rightarrow & H(X,Y,Z) - H(Y,Z) - H(X,Y) + H(Y) \leq \\ & \leq \frac{1}{\ln 2}\sum_{x,y,z} p(x,y,z)\left(\frac{p(x|y)}{p(x|y,z)} - 1\right) = \\ & = \frac{1}{\ln 2}\sum_{x,y,z}(p(y,z)p(x|y) - p(x,y,z)) = \\ & = \frac{1}{\ln 2}\sum_{x,y} p(x|y)p(y) - 1 = \frac{1}{\ln 2}\sum_{x,y} p(x,y) - 1 = 0\end{aligned} \tag{1.63}$$

where I used that $p(x|y,z) = p(x,y,z)/p(y,z)$. This concludes the proof. Notice that the condition of equality is satisfied if $p(x|y,z) = p(x,y)$, which is the definition for a Markov chain $Z \to Y \to X$. □

Lastly there is the chain rule for conditional entropies

$$H(X_1,\ldots,X_n|Y) = \sum_{i=1}^{n} H(X_i|Y,X_1,\ldots,X_{i-1}) \tag{1.64}$$

Proof. For $n = 2$ we have

$$H(X_1,X_2|Y) = H(X_1,X_2,Y) - H(Y) = H(X_2|Y,X_1) + H(X_1|Y) \tag{1.65}$$

adding and subtracting $H(X_1, Y)$. Let us go for an induction proof and assume that the inductive hypotesis holds, then

$$\begin{aligned} H(X_1, \ldots, X_{n+1}|Y) &= H(X_2, \ldots, X_{n+1}|Y, X_1) + H(X_1|Y) = \\ &= \sum_{i=2}^{n+1} H(X_i|Y, \ldots, X_{i-1}) + H(X_1|Y) = \\ &= \sum_{i=1}^{n+1} H(X_i|Y, X_1, \ldots, X_{i-1}) \end{aligned} \tag{1.66}$$

□

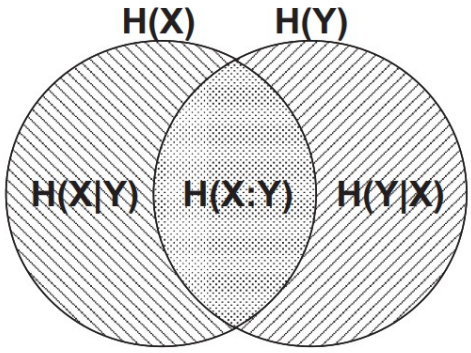

Figure 1.4: In this image is a clear and simple diagram showing the relations between the kind of entropies we saw until now. ([30], p.508)

1.4.2 Quantum Entropies

To work in the quantum mechanics framework we have to go from probability distributions to density matrices. To do so we can define the von Neumann entropy.

Definition 1.4.5 (**von Neumann entropy**).

$$H(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log \lambda_i$$

where λ_i are the eigenvalues of ρ. Similarly to the classical case this quantity represents the uncertainty that we have regarding the system represented by ρ. As before it's useful to introduce the quantum relative entropy as well.

Definition 1.4.6 (**Quantum Relative entropy**)**.**

$$H(\rho||\sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma)$$

and $H(\rho||\sigma) = +\infty$ if $\mathrm{supp}\rho \cap \ker\sigma \neq 0$, i.e. if there is an overlap between the kernel of σ (the space spanned by the eigenvectors with eigenvalue 0) and the support of ρ (the space spanned by eigenvectors with eigenvalue different from 0).

1.4.2.1 Mathematical Properties

As expected the von Neumann entropy and the quantum relative entropy are non-negative

$$H(\rho) \geq 0 \tag{1.67}$$

Proof. Since $\rho \geq 0$ and $\sum_i \lambda_i = 1$ the proof is trivial. The equality here is fulfilled $\iff$ ρ is a pure state. We can see this simply:

- $\Rightarrow$ If the entropy vanishes we have $\lambda_j = 0 \vee 1 \, \forall j$, but $\sum_j \lambda_j = 1$ so only one of the eigenvalues can be 1 and all the other are 0. This is the definition of pure state.
- $\Leftarrow$ If the state is pure only one eigenvalue is different from 0, and from $\sum_j \lambda_j = 1$ it will be equal to 1. Thus $H(\rho) = -\log(1) = 0$.

□

Another property is that of *additivity*:

$$H(\rho \otimes \sigma) = H(\rho) + H(\sigma) \tag{1.68}$$

Proof. From the definition we have

$$H(\rho \otimes \sigma) = -\mathrm{Tr}[\rho \otimes \sigma \log(\rho \otimes \sigma)] \tag{1.69}$$

and we need to manipulate the logarithm to continue, and expliciting again

$$\rho = \sum_i \lambda_i \left|i\right\rangle \left\langle i\right| \ , \ \sigma = \sum_j \mu_j \left|j\right\rangle \left\langle j\right| \tag{1.70}$$

makes the calculations simpler

$$\begin{aligned}
&\log(\rho\otimes\sigma) = \sum_{i,j}\log(\lambda_i\mu_j)\,|i\rangle\langle i|\otimes|j\rangle\langle j| = \\
&= \sum_{i,j}\log\lambda_i\,|i\rangle\langle i|\otimes|j\rangle\langle j| + \sum_{i,j}\log\lambda_i\,|i\rangle\langle i|\otimes|j\rangle\langle j| = \\
&= \sum_{i}\log\lambda_i\,|i\rangle\langle i|\otimes\mathbb{1}_\sigma + \mathbb{1}_\rho\otimes\sum_{j}\log\mu_j\,|j\rangle\langle j| = \\
&= \log\rho\otimes\mathbb{1}_\sigma + \mathbb{1}_\rho\otimes\log\sigma
\end{aligned} \tag{1.71}$$

thus

$$\begin{aligned}
H(\rho\otimes\sigma) &= -\mathrm{Tr}[\rho\otimes\sigma(\log\rho\otimes\mathbb{1}_\sigma + \mathbb{1}_\rho\otimes\log\sigma)] = \\
&= -\mathrm{Tr}[\rho\log\rho\otimes\sigma + \rho\otimes\sigma\log\sigma] = \\
&= -\mathrm{Tr}[\rho\log\rho]\cdot\mathrm{Tr}[\sigma] - \mathrm{Tr}[\rho]\cdot\mathrm{Tr}[\sigma\log\sigma] = \\
&= H(\rho) + H(\sigma)
\end{aligned} \tag{1.72}$$

where I used the property of the trace that $\mathrm{Tr}[A\otimes B]$=$\mathrm{Tr}[A]\cdot\mathrm{Tr}[B]$ and the fact that $\mathrm{Tr}[\rho]$ =$\mathrm{Tr}[\sigma] = 1$ from the properties of density matrices. □

A slightly less articulated proof is that of the concavity of the von Neumann entropy. Given $\{p_j\}$, $\{\rho_j\}$ with $\sum_j p_j = 1$, $p_j \geq 0$ and $\rho_j \in S(\mathcal{H})\ \forall j$

$$H\left(\sum_j p_j\rho_j\right) \geq \sum_j p_j H(\rho_j) \tag{1.73}$$

Proof. Taking $\lambda_{i,j}$ as the j-th eigenvalue of the i-th density matrix we have

$$H\left(\sum_j p_j\rho_j\right) = -\sum_{i,j} p_i\lambda_{i,j}\log p_i\lambda_{i,j} \tag{1.74}$$

from Def. 1.4.5, thus

$$\begin{aligned}
&H\left(\sum_j p_j\rho_j\right) = -\sum_{i,j} p_i\lambda_{i,j}(\log p_i + \log\lambda_{i,j}) = \\
&= -\sum_i p_i\log p_i\sum_j\lambda_{i,j} - \sum_i p_i\sum_j\lambda_{i,j}\log\lambda_{i,j} = \\
&= S(p) + \sum_i p_i H(\rho_i)
\end{aligned} \tag{1.75}$$

where I used $\sum_j \lambda_{i,j} = 1$, $\forall i$. In the last line $S(p)$ is the Shannon entropy of the probability distribution p, and since $S(p) \geq 0$ this proves the concavity of the von Neumann entropy. □

Next we can prove the positivity of the relative entropy

$$H(\rho||\sigma) \geq 0 \tag{1.76}$$

Proof. Rewriting explicitly the density operators as

$$\rho = \sum_i \lambda_i |i\rangle \langle i| \ , \ \sigma = \sum_j \mu_j |j\rangle \langle j| \tag{1.77}$$

the entropy can be rewritten as

$$H(\rho||\sigma) = \sum_i \lambda_i \log \lambda_i - \sum_i \langle i| \rho \log \sigma |i\rangle \tag{1.78}$$

and, since $\langle i| \rho = \lambda_i \langle i|$ and

$$\langle i| \log \sigma |i\rangle = \sum_j \log \mu_j P_{ij} \tag{1.79}$$

with $P_{ij} = \langle i|j\rangle \langle j|i\rangle \geq 0, \sum_i P_{ij} = \sum_j P_{ij} = 1$, then

$$\begin{aligned} H(\rho||\sigma) &= \sum_i \lambda_i \left(\log \lambda_i - \sum_j Pij \log \mu_j \right) \geq \sum_i \lambda_i (\log \lambda_i - \log ri) = \\ &= \sum_i \lambda_i \log \frac{\lambda_i}{r_i} \end{aligned} \tag{1.80}$$

where the inequality is due to the concavity of log, taking $r_i = \sum_j P_{ij} \mu_j$. This is the same form of the classical relative entropy, thus (1.50) concludes the proof. □

As in the classical case we have that in a d-dimensional Hilbert space the entropy is upper bounded by $H(\rho) \leq \log d$, and the proof goes in the same way, having that $0 \leq H(\rho||\mathbb{1}/d) = -H(\rho) + \log d$, with equality $\iff \rho = \mathbb{1}/d$.

1.4.2.2 Joint and Conditional Entropy

Analogously to what we did for classical entropies we can define a joint quantum entropy

Definition 1.4.7 (**Joint quantum entropy**)**.** Given $\rho_{AB} \in \mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ the joint entropy is

$$H(\rho_{AB}) = -\mathrm{Tr}[\rho_{AB} \log \rho_{AB}] \tag{1.81}$$

Here we can start to notice some differences, in fact, while for Shannon entropy it held that $H(X) \leq H(X, Y)$ given two random variables X and Y, in the case of von Neumann entropy it does not. Take the entangled state $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ with the first qubit in $\mathcal{H}_A$ and the second in $\mathcal{H}_B$, in this case the joint entropy $H(A, B)$ is 0 since this state is pure, but doing the partial trace over the space of the second qubit we get $H(A) = \mathrm{Tr}_B[S(A, B)] = 1$ since the state on this space is the completely mixed state. This introduces a peculiar consequence for the conditional quantum entropy

$$H(A|B) = H(A, B) - H(B) \tag{1.82}$$

which is that $H(A|B)$ **can** be negative, and it is for all (pure) entangled states[9].

As in the classical case conditioning reduces the entropy, meaning that for the von Neumann entropy as well it holds that

$$H(A|B, C) \leq H(A|B) \tag{1.83}$$

and the proof is the same that was used in (1.58).

1.4.2.3 (Smooth) Min- and (Smooth) Max-Entropy

In this paragraph I will present what are known as *one-shot entropies*, following mainly [48] (Chapter 4) and [55] (Chapter 3.2.6). The name comes from the fact that these specific entropies allow to quantify information in non-IID cases, and even though the project in question takes IID assumptions, these one-shot entropies will be useful for a range of tasks. To understand where one-shot entropies come from I will first introduce the *classical Rényi α-entropy*

Definition 1.4.8 (**Rényi α-entropy**)**.** Taken a discrete deistribution P_X over a set $\mathcal{X}$ the Rényi α-entropy is

$$H_\alpha(X)_P = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X(x)^\alpha \tag{1.84}$$

with $\alpha \in (0, 1) \cup (1, +\infty)$

[9] For mixed entangled states the conditioned entropy can still be negative but not necessarily.

which can be generalized to the quantum setting introduce the density matrix formalism

Definition 1.4.9 (**Quantum Rényi α-entropy**)**.** Taken a (normalized) state $\rho \in S(\mathcal{H}_A)$ the quantum Rényi α-entropy is

$$H_\alpha(A)_\rho = \frac{1}{1-\alpha} \log \mathrm{Tr}[\rho_A^\alpha] \tag{1.85}$$

In this case we'll extend the domain of α to include 0,1 and $+\infty$ as well, just taking the limit of the definition

$$H_\infty = -\log ||A||_\infty \,,\; H_0 = \log \mathrm{rank}[\rho_A] \,,\; H_1 \equiv H_{VonNeumann} \tag{1.86}$$

These entropies are monotonically decreasing in the value of α, thus $H_\infty(A)_\rho \leq \cdots \leq H_1(A)_\rho \leq \cdots \leq H_0(A)_\rho \; \forall \rho \in S(\mathcal{H}_A)$, which is a first way to see why we will call H_∞ the min-entropy and $H_{\frac{1}{2}}$ the max-entropy[10].

Starting from the min-entropy, one of the definitions that are commonly found is

Definition 1.4.10 (**Min-Entropy**)**.** Let $\rho_{AB} \in S_\leq(\mathcal{H}_{AB})$ with $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ and $S_\leq(\mathcal{H})$ be the space of subnormalized states in an Hilbert space. The min-entropy of A conditioned on B is

$$H_{min}(A|B)_\rho = \max_\sigma \sup \left[\lambda \in \mathbb{R} : \rho_{AB} \leq 2^{-\lambda} \mathbb{1}_A \otimes \sigma_B\right] \tag{1.87}$$

where the maximum is taken over all states $\sigma \in S_\leq(\mathcal{H}_B)$.

This definition is practically convenient, as it trivialize the formulation of min-entropy as an *SDP* (semi-definite program)[11], another (equivalent) definition is

$$\begin{aligned} H_{min}(A|B) &= \sup_{\sigma_B} H_{min}(\rho_{AB}|\sigma_B) = \\ &= -\log \left[\min_{\sigma_B} \big(\min \left[\lambda \in \mathbb{R} : \rho_{AB} \leq \lambda \cdot \mathbb{1}_A \otimes \sigma_B\right] \big)\right] \end{aligned} \tag{1.88}$$

since

$$H_{min}(\rho_{AB}|\sigma_B) = -\log \min \left[\lambda \in \mathbb{R} : \rho_{AB} \leq \lambda \cdot \mathbb{1}_A \otimes \sigma_B\right] \tag{1.89}$$

[10] Furthermore we have that H_∞ approximates well the Rényi entropies with $\alpha > 1$, while $H_{\frac{1}{2}}$ approximates well the entropies with $\alpha < 1$, for details see [48] Chapter 4.1.

[11] In Chapter 2.3.1 of [48] and in Watrous' lecture notes [52] the topic is studied more in detail.

Moving away from technical definitions for a moment we can interpret this as, taken a classical–quantum states, the conditional min-entropy characterizes the amount of uniform randomness that we can extract from a classical random variable that is correlated with a quantum system such that the result is independent of the quantum system ([55], [25], [34]).

Following is the max-entropy

Definition 1.4.11 (**Max-Entropy**)**.** Let $\rho_{AB} \in S_{\leq}(\mathcal{H}_{AB})$. The max-entropy of A conditioned on B is

$$H_{max}(A|B)_\rho = \max_\sigma \log \left|\left|\sqrt{\rho_{AB}}\sqrt{\mathbb{1}_A \otimes \sigma_B}\right|\right|_1^2 \tag{1.90}$$

where the maximum is taken over all states $\sigma \in S_{\leq}(\mathcal{H}_B)$.

The max-entropy can be interpreted in two ways: the first one is as a the distance of ρ from a state describing an observer B which is ignorant of A ([48]), the second is as a quantifier of the size of the system that A can be compressed to, such that the original system can be recovered, given access to the quantum system B ([55]). The next step is easy, because to define formally the *smooth* min-/max-entropy the only thing we need to do is to do a smoothing:

Definition 1.4.12 (**Smooth Entropies**)**.** Defined an ε-ball around a state ρ as

$$\mathcal{B}^\varepsilon(\rho) = \{\tau : \tau \in S_{\leq}(\mathcal{H}_{AB}), P(\rho, \tau) \leq \varepsilon\} \tag{1.91}$$

where $P(\rho, \tau) = \sqrt{1 - F(\rho, \tau)}$ is the purified distance between two states, the smooth min-/max-entropy are

$$H_{min}^\varepsilon(A|B)_\rho = \max_{\rho' \in \mathcal{B}^\varepsilon(\rho)} H_{min}(A|B)_{\rho'} \tag{1.92}$$

$$H_{max}^\varepsilon(A|B)_\rho = \min_{\rho' \in \mathcal{B}^\varepsilon(\rho)} H_{max}(A|B)_{\rho'} \tag{1.93}$$

As in (1.88) it's possible to reformulate (1.92), (1.93) in terms of supremum and infimum over states, the definition is equivalent and is present in Chapter 3.2.1 of [34]. Operationally it's useful for us to consider the smooth min-entropy as directly related to the number of bits we can extract from A that, from the point of view of B, are ε-distant from the uniform distribution[12]. Other practical examples that help understand these (unfamiliar) quantities are presented in [25].

[12]This result is known as the *Leftover Hash Lemma* and will be explored into detail further in this work (Section 2.2.2).

Lastly an important relationship is that between the min-entropy and the guessing probability. Indeed we have that the min-entropy $H_{min}(A|B)$ can be interpreted as the probability that an observer with access to B guesses A, and this is formalized as

$$p_{guess}(A|B) = 2^{-H_{min}(A|B)}\,,\ H_{min}(A|B) = -\log p_{guess} \tag{1.94}$$

1.4.3 Entropy relations

Now that we have all the tools necessary I will present some of the common and useful relations between entropies.

1.4.3.1 Chain Rules

While I consider important going through the proofs at least once to truly understand what we are working with, I will not present the proofs in this chapter, since they are non-trivial and require introducing more entropic quantities and theorems than necessary, however all the claims done here are proven in [34].

Given the monotonicity of Rényi entropies one first property is that the min-entropy cannot be larger than the max-entropy

Lem 1.4.1. *Let $\rho_{AB} \in S_{\leq}(\mathcal{H}_{AB})$ and $\sigma_B \in S(\mathcal{H}_{\mathcal{B}})$, then*

$$H_{min}(\rho_{AB}|\sigma_B) + \log \mathrm{Tr}[\rho_{AB}] \leq H_{max}(\rho_{AB}|\sigma_B) \tag{1.95}$$

Another property, as for the von Neumann entropy, is that for a state consisting of two (or more) fully independent parts the min- and max-entropy are additive

$$H_{min}(\rho_{AB} \otimes \rho_{A'B'}|\sigma_B \otimes \sigma_{B'}) = H_{min}(\rho_{AB}|\sigma_B) + H_{min}(\rho_{A'B'}|\sigma_{B'}) \tag{1.96}$$
$$H_{max}(\rho_{AB} \otimes \rho_{A'B'}|\sigma_B \otimes \sigma_{B'}) = H_{max}(\rho_{AB}|\sigma_B) + H_{max}(\rho_{A'B'}|\sigma_{B'}) \tag{1.97}$$

Unsurprisingly these one-shot entropies also share the property of strong subadditivity with the von Neumann entropy

$$H_{min}(\rho_{ABC}|\sigma_{BC}) \leq H_{min}(\rho_{AB}|\sigma_B) \tag{1.98}$$
$$H_{max}(\rho_{ABC}|\sigma_{BC}) \leq H_{max}(\rho_{AB}|\sigma_B) \tag{1.99}$$

Introducing classical information, i.e. classical subsystems, there are two more useful results

$$H_{min}(\rho_{XC}|\rho_C) \geq 0 \tag{1.100}$$

where X is a classical system[13], and the same holds for the max-entropy.

The other result related to classical-quantum systems is

$$H_{min}(\rho_{ABZ}|\sigma_{BZ}) = \inf_{z\in\mathcal{Z}} H_{min}(\rho^z_{AB}|\sigma^z_B) \tag{1.101}$$

$$H_{max}(\rho_{ABZ}|\sigma_{BZ}) = \log \sum_{z\in\mathcal{Z}} 2^{H_{max}(\rho^z_{AB}|\sigma^z_B)} \tag{1.102}$$

where ρ_{ABZ} and σ_{BZ} are classical with respect to the orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$, meaning that exist a family $\{\rho^z_{AB}\}_{z\in\mathcal{Z}}$ and $\{\sigma^z_B\}_{z\in\mathcal{Z}}$ called *(non-normalized) conditional operators*, such that

$$\rho_{AB} = \sum_{z\in\mathcal{Z}} \rho^z_{AB} \otimes |z\rangle\langle z| \ , \ \sigma_B = \sum_{z\in\mathcal{Z}}^{z} \sigma_B \otimes |z\rangle\langle z| \tag{1.103}$$

What is more properly known as the chain rule for min-entropy is the extension of the result for the von Neumann entropy $H(A,B|C) \leq H(A|B,C) + H(B)$, derived from the definition $H(A,B|C) = H(A|B,C) + H(B|C)$ and (1.58)

$$H_{min}(\rho_{ABC}|\sigma_C) \leq H_{min}(\rho_{ABC}|\sigma_B \otimes \sigma_C) + H_{max}(\rho_B) \tag{1.104}$$

where σ_B is the fully mixed state on the support of ρ_B.

The smooth min-entropy is *superadditive*: let $\rho_{AB} \in S_{\leq}(\mathcal{H}_{AB}), \sigma_B \in S_{\leq}(\mathcal{H}_B)$ and let $\varepsilon, \varepsilon' \geq 0$, then

$$H^{\varepsilon+\varepsilon'}_{min}(\rho_{AB} \otimes \rho_{A'B'}|\sigma_B \otimes \sigma_{B'}) \geq H^{\varepsilon}_{min}(\rho_{AB}|\sigma_B) + H^{\varepsilon'}_{min}(\rho_{A'B'}|\sigma_{B'}) \tag{1.105}$$

as well as *strongly subadditive*: let $\rho_{ABC} \in S_{\leq}(\mathcal{H}_{ABC}), \sigma_{BC} \in S_{\leq}(\mathcal{H}_{BC})$ and let $\varepsilon \geq 0$, then

$$H^{\varepsilon}_{min}(\rho_{ABC}|\sigma_{BC}) \leq H^{\varepsilon}_{min}(\rho_{AB}|\sigma_B) \tag{1.106}$$

[13]If all systems are quantum, and in particular if all systems are entangled the entropy can be negative.

Analogously to the non-smoothed case of the min-entropy we have that

$$H^{\varepsilon}_{min}(\rho_{ABZ}|\sigma_{BZ}) \geq \inf_{z\in\mathcal{Z}} H^{\varepsilon}_{min}(\rho^{z}_{AB}|\sigma^{z}_{AB}) \tag{1.107}$$

with states classical with respect to the orthonormal basis $\{|z\rangle\}_{z\in\mathcal{Z}}$, and the chain rule is given by

$$H^{\varepsilon}_{min}(\rho_{ABC}|\sigma_C) \leq H^{\varepsilon}_{min}(\rho_{ABC}|\sigma_B \otimes \sigma_C) + H_{max}(\rho_B) \tag{1.108}$$

where again σ_B is the fully mixed state on the support of ρ_B.

1.4.3.2 Entropy Accumulation Theorem

One important theorem is the Entropy Accumulation Theorem (EAT), formulated in [15] and then improved by [14]. The EAT was the basis to many other important theorems, such as the Generalized EAT [27] or the Marginal EAT [3].
Firstly I will present the EAT, then analyze the quantities appearing, and later I will discuss the meaning of this result. However, there are two conditions that need to be true for the EAT to hold: for $i \in \{1,\dots,n\}$ let $\mathcal{M}_i$ be CPTP maps[14] from R_{i-1} to $X_iA_iB_iR_i$, where A_i and B_i are the parties, of which only A_i need to be finite dimensional, X_i is the classical value determined by both A_i and B_i, and R_i is the system under analysis after i iterations of some protocol. Then the EAT holds for:

1. states of the form

$$\rho_{A_1^nB_1^nX_1^nE} = (\mathcal{M}_1 \circ \cdots \circ \mathcal{M}_n \circ \mathbb{1}_E)(\rho^0_{R_0E}) \tag{1.109}$$

   where $\rho^0_{R_0E}$ is a density operator on the initial system R_0 and an arbitrary system E.

2. Markov chains[15] which $\forall i \in \{1,\dots,n\}$ have

$$A_1^{i-1} \longleftrightarrow B_1^{i-1}E \longleftrightarrow B_i \tag{1.110}$$

[14]CPTP stands for completely positive and trace preserving. Taken a map $\mathcal{E}$ it is TP $\iff$ Trρ=Tr$[\mathcal{E}(\rho)]$ and it is CP $\iff$ $\mathcal{E}\otimes\mathbb{1}_{anc}$ is positive $\forall n \in \mathbb{N}$, where $\mathbb{1}_{anc}$ is the identity on an ancillary system.

[15]A Markov chain is (qualitatively) a chain of events where the future state only depends on the present state of the chain, i.e. future states are conditionally independent of past states when the present state is known.

Theorem 1.4.2 (**Entropy Accumulation Theorem**). *Let $\mathcal{M}_1, \dots, \mathcal{M}_n$ and $\rho_{A_1^n B_1^n X_1^n E}$ be such that* (1.109) *and the Markov conditions* (1.110) *hold, let $h \in \mathbb{R}$ and let f be an affine min-tradeoff function for $\mathcal{M}_1 \dots, \mathcal{M}_n$, finally let $\varepsilon \in (0,1)$. Then for any event $\Omega \subseteq \mathcal{X}^n$ that implies $f(\mathbf{freq}(X_1^n)) \geq h$ we have*

$$H_{min}^{\varepsilon}(A_1^n|B_1^n, E)_{\rho|\Omega} > nh - c\sqrt{n} \tag{1.111}$$

for $c = 2\big[\log(1+2d_A) + \lceil ||\nabla f||_\infty \rceil\big]\sqrt{1 - 2\log(\varepsilon\rho[\Omega])}$, where d_A is the maximum dimension of A. Replacing f for an affine max-tradeoff function yields

$$H_{max}^{\varepsilon}(A_1^n|B_1^n, E)_{\rho|\Omega} < nh - c\sqrt{n} \tag{1.112}$$

if Ω implies $f(\mathbf{freq}(X_1^n)) \leq h$.

It might be easier to understand studying all the pieces one by one:

- f is a real affine function ($f(x) = ax + b$ with $a, b \in \mathbb{R}$) on the set of probability distributions over an alphabet $\mathcal{X}$[16], so that the gradient (∇f) will be constant, a technical constraint to keep the convexity of the set of probability distributions ([14]).

- Ω is an event defined by a subset of $\mathcal{X}^n$, where $\mathcal{X}$ is the alphabet (the set of possible outcomes, disregarding whether they have probability 0 or non-null) for the distribution, and n represents either the number of parties which get one value of the distribution, or the number of outcomes a single party gets from the distribution. The probability that the event Ω happens is

  $$\rho[\Omega] = \sum_{x_1^n \in \Omega} \mathrm{Tr}[\rho_{A_1^n B_1^n E, x_1^n}] \tag{1.113}$$

  where x_1^n are the actual values of the distribution, A_1 and E are the parties involved.

- $\mathbf{freq}(X_1^n)$ is the distribution on $\mathcal{X}$ defined by

  $$\mathbf{freq}(X_1^n)(x) = \frac{|\{i \in \{1, \dots, n\} : X_i = x\}|}{n} \tag{1.114}$$

- In case of infinite-dimensional A the quantity d_A can be substituted with appropriate entropic quantities, as we will see in the QAEP.

- h will be used to bound the von Neumann entropy.

[16] The exact definition and more details are given in Chapter 4 of [15].

The two conditions (1.110) and (1.109) put some constraints on the theorem, in particular the second assumption is weaker than IID conditions, but stronger than non-IID conditions. This is however addressed in the formulation of the GEAT [27], where the assumptions are No-Signaling (NS) (weaker than non-IID) rather than Markovian.
Putting aside the technical parts of the EAT, the relevance of this result is to be found in the fact that, under (fairly) weak assumptions, it is possible to relate the entropy of a bitstring (or a set of outcomes) conditioned on classical and quantum information, to the entropy of each bit (conditioned on the same quantities but this time **per bit**), with a correction term which takes into account finite-length scenarios and that grows as $O(\sqrt{n})$.

1.4.3.3 Quantum Asymptotic Equipartition Property

Similar to the EAT we have a more specific result, which is the Quantum Asymptotic Equipartition Property, or QAEP. This was first shown in [49], where the authors find a fully quantum generalization of the classical information theory Asymptotic Equipartition Property. In a much less cumbersome way if compared to the EAT, the QAEP simply is

Theorem 1.4.3 (**Quantum Asymptotic Equipartition Property**). *Let $\mathcal{H}_A, \mathcal{H}_B$ be finite-dimensional Hilbert spaces, ρ_{AB} a bipartite state on their tensor product and $n \in \mathbb{N}$ such that $\rho_{AB}^{\otimes n}$ is an IID state on $(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes n}$, then*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_{\rho} \tag{1.115}$$

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{max}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_{\rho} \tag{1.116}$$

In case of finite-length we have

$$H_{min}^{\varepsilon}(A^n|B^n)_{\rho^{\otimes n}} \geq n \cdot H(A|B)_{\rho} - \sqrt{n} \cdot \delta(\varepsilon, \eta) \tag{1.117}$$

where $\delta(\varepsilon, \eta) = 4 \log \eta \sqrt{\log(2/\varepsilon^2)}$ for $n \geq \frac{8}{5} \log \frac{2}{\varepsilon^2}$ and $\eta = \Upsilon(A|B)_{\rho|\rho}$

For completeness we have that $\Upsilon(A|B)_{\rho|\sigma}$ is a parameter defined as the α-entropy convergence parameter, defined as

$$\Upsilon(A|B)_{\rho|\sigma} = 2^{-\frac{1}{2} H_{3/2}(A|B)_{\rho|\sigma}} + 2^{\frac{1}{2} H_{1/2}(A|B)_{\rho|\sigma}} + 1 \tag{1.118}$$

of which η is a special case with $\sigma = \rho$. Using

$$H_{max}(A|B) \geq H_{3/2}(A|B)$$
$$H_{min}(A|B) \leq H_{1/2}(A|B) \tag{1.119}$$

we have the simplified parameter

$$\eta = \Upsilon(A|B)_{\rho|\rho} \leq \sqrt{2^{-H_{max}(A|B)_\rho}} + \sqrt{2^{H_{min}(A|B)_\rho}} + 1 \tag{1.120}$$

This quantity in the general case ($\sigma \neq \rho$) represents how quickly the α-entropies converge to the von Neumann entropy as a consequence of the QAEP: a smaller η implies a faster convergence and thus a smaller correction. To have an idea of the order of magnitude of this quantity we can take the values for $H^{\varepsilon}_{min}(A|B)$ and $H^{\varepsilon}_{max}(A|B)$ that maximize η, so respectively 0 and 1, so the upper bound for $\log \eta$ is ≈ 1.77.
Again we have a very important tool that connects the (smooth min-) entropy accumulated over multiple rounds to the entropy of the bits of each rounds, taking into account finite-length corrections.

Chapter 2

Quantum Cryptography

This section is dedicated to introducing concepts related to quantum cryptography, following mainly [21] and [32]. Cryptography has ancient roots, and classical cryptography consists mostly of combining a message with a key through an algorithm, or cipher, in such a way that only authorized users can decrypt it. There are two kinds of classical encryptions, *asymmetric* and *symmetric* cryptosystems:

- In an asymmetric (or public-key) cryptosystem A and B use two different keys for encryption and decryption; B creates a private key which is kept secret, and using that key creates a public key which is shared with whomever wants to share information. Then A uses this public key to encrypt the message and send it to B, who will decrypt it using the secret key. An example is the *RSA* cryptosystem ([38]), which is the basis of many encryptions over the internet. Qualitatively it works like this: B uses some algorithm to choose two "very big" prime numbers p, q multiply them ($n = pq$) and find the Euler's totient of the two ($\varphi(p,q) = (p-1)(q-1)$), then B picks another large integer which is coprime with $\varphi(p,q)$ and finally gets $e = \frac{1}{d} \mathrm{mod} \varphi(p,q)$. Now the public key is the pair (n, e) and the message M can be ciphered by A as $C = M^e \mathrm{mod} n$, and deciphered by B as $M = C^d \mathrm{mod} n$. Functions such as n are called one-way functions because it is easy to calculate $f(x)$ given x but it's very hard to calculate x knowing only $f(x)$. While asymmetric encryption is largely used today, it lacks a real proof of security since it is really hard to show that a classical algorithm which efficiently factorizes big numbers into their primes does not exist.

- The symmetric encryption only uses one key for both encryption and decryption, the most famous example is the *one-time pad* (OTP). In the OTP protocol A and B share a secret key, which must be physically exchanged by secure means, that can be used to mask the message. In this way, using the fact that $k \oplus k = 1$ for bits, A can send B the message m as $m' = m \oplus k$

and B can decrypt it summing again the key $m' \oplus k = m$. The disadvantages are that the key must be at least as long as the message and can be used only once per message, to avoid giving extra information to external eavesdropper.

With the advent of quantum computing new protocols based on the efficiency of quantum computers were proposed (a famous example being Shor's algorithm, which can break the RSA cryptography in polynomial time, [43]) and thus we developed the need of encryption models secure against quantum attacks. It is in this context that the idea of quantum cryptography is born, a set of encryption protocols and methods which are based on the unbreakability of the laws of physics, rather than the absence of strong enough computers. For reasons of practicality and urgency post-quantum, or quantum-proof, classical encryption was born as well, to allow for security against quantum attacks without having to wait for a true development of quantum computers.

2.1 Quantum Key Distribution

Quantum key distribution (QKD) looks like the best solution for encryption, but it has the downside of requiring to find a costly equilibrium between security and efficiency: the less trust I put in my devices the lower is the secret key rate. In this section I will present some of the introductory QKD protocols and some more advanced versions, all in the context of discrete-variable QKD (DVQKD), leaving the treatment of continuous-variable QKD (CVQKD) to others more expert than me.

2.1.1 Prepare-and-Measure and Entanglement-based protocols

The first type of QKD protocols in this section are the prepare-and-measure protocols, where the information is encoded in quantum properties of whatever gets exchanged, for example in the polarization of photons. This allows to spot and eliminate any run of the protocol that is wrong, whether it is because of noise or because of adversaries. A convention is to call the transmitter A, who prepares optical signals, and the receiver B, who measures them to retrieve the encoded information.

The BB84 protocol was the first ever proposed QKD protocol ([6][1]) and, using again the convention that A is the transmitter and B the receiver, following are the main steps:

1. A prepares 4 random states in two different bases, usually the $\{|0\rangle, |1\rangle\} \equiv Z$ basis and $\{|+\rangle, |-\rangle\}X$ basis (presented in (1.5)[2]), and associates binary values with non-orthogonal pairs (0 if $|0\rangle$ or $|+\rangle$, 1 otherwise) so that the non-orthogonality of the states guarantees the impossibility of cloning or measuring with perfect fidelity by an eavesdropper.

2. A sends B the states and B measures them using a random basis between X and Z. If B randomly chooses the same basis as A for a particular state they measure the same bit, otherwise the bit will be random.

3. A and B announce publicly which basis they used to measure each state and the bits associated to different bases are discarded.

4. A and B agree on a subset of bits that they publicly announce in order to compare them, if the bits agree (up to a certain error) they are discarded and the protocol continues, otherwise the protocol is aborted and restarts from step 1.

5. The remaining bits are the final shared key, which can be used for example in a one-time pad scenario.

In [32] Section B.1 is also proved that this strategy maximizes the information recovery for A and B, while an eavesdroppes E can only reach half of their recovery rate.

A better[3] version of BB84 is known as the *SARG04* and was proposed in [40], where the authors found a way to make BB84 more robust against a specific kind of attack. The protocol shares the same steps as BB84 but the step 3. is substituted by:

5. B announces their bases, A announces a pair of non-orthogonal states, such as $\{|0\rangle, |+\rangle\}$ or $\{|1\rangle, |-\rangle\}$ (both pairs have an overlap of $1/\sqrt{2}$ between the two states), where $\{|0\rangle, |1\rangle\}$ encode a 0 and $\{|+\rangle, |-\rangle\}$ a 1.

[1] This paper is a transcription of the contribution of the authors at the "International Conference on Computers, Systems & Signal-Processing" held in Bangalore, India between the 9th and the 12th of December 1984.

[2] The minimum number of states necessary is 2, and on this is based for example the B92 protocol, while other protocols, such as the six-state protocol, introduce the Y basis as well, which makes the eavesdropper easier to spot.

[3] Less vulnerable to photon number splitting attacks, see Section 2.1.2.

For example, imagine A sends $|0\rangle$ and announces the set $\{|0\rangle, |+\rangle\}$, if B measured σ_X then it's not possible to infer which of the two states was originally since both can give a $+1$, and the same goes if B measures in σ_Z and gets a $+1$ as the outcome, and that bit must be thrown away. However if B gets a -1 as the outcome, then the original state is without a doubt $|0\rangle$ and B adds a 0 to the secret key. This is advantageous because announcing pairs of non-orthogonal states reduces the information available to E in case of an attack, at the price of a slightly less efficient protocol with a lower key rate.

Unlike prepare-and-measure protocols, which rely on the no-cloning theorem (see Section 3.2.2 for a brief explanation) and the impossibility of distinguishing perfectly two non-orthogonal states, one could use nonlocality to create secure communication, where entangled pairs of particles are used to check Bell-like tests to rule out the presence of eavesdroppers. This new approach to QKD was first developed by Arthur Ekert in 1991, when he proposed the *E91* protocol ([17]). Consider the scenario where a single source emits pairs of entangled particles in the Bell state $|\psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$, then the protocol is:

1. A and B measure their own particle, A choosing a random basis between $\{Z\}, \{(X+Z)/\sqrt{2}\}, \{X\}$ and B choosing randomly from $\{(X+Z)\sqrt{2}\}, \{X\}$, $\{(X-Z)/\sqrt{2}\}$.

2. A and B discuss publicly the bases they used for their measurements and use the mismatched rounds to check the presence of E.

To check the presence of adversaries, they study a CHSH quantity and if the Bell-like inequality holds then either the photons were not entangled or there is an eavesdropper, so the protocol must be aborted, while in case of a violation they can process the remaining bits to create a secret key.

Just a year later however a similar protocol, the *BBM92*, was published in [7] with the plus of using just two bases, similarly to BB84, and without the need of doing CHSH tests. The similarities do not end here, in fact BBM92 is an entanglement-based representation of BB84 since, it's not possible for B to determine whether they received half of an entangled pair or if A prepared a state and then measured it. The idea of bridging prepare-and-measure and entanglement-based protocols as representations of one another is a useful theoretical tool that allows to study the security and correctness of only one protocol out of a set of all equivalent protocols.

2.1.2 Quantum hacking

A QKD protocol is never perfect when put into practice in reality, and any of the imperfections and ideal assumptions that one takes during the formulation of a protocol might not hold, giving rise to some *side-channels*. When this happens an adversary can implement some attacks which are aimed at gaining more information, and *quantum hacking* is the term that encompasses all attacks of this kind. Now I will very briefly list some of the most common types of attacks that an adversary can perform and how the parties can mitigate the damage.

- **Photon number splitting (PNS) and intensity-based attacks** Many security proofs, such as the one for the BB84 protocol, are based on the assumption that the sources are single-photon emitters, an assumption which is almost never true in reality, where the sources use highly dampened lasers emitting multiple photons. What E could do is intervene in the middle of the line and using some beam-splitters reroute all but one photons and store them in a quantum memory[4], just waiting for the classical post-processing to be over. At that point E can perform a collective measurement on the stolen photons and recover the exact key bit by bit. This attack is undetectable because E does not interact with the photons until the classical communication is over, so they parties do not know that some extra measurement took place, and it is for this reason that a protocol such as the SARG04 in Section 2.1.1 exist, addressing this side-channel and making it secure once again.

- **Man-in-the-middle and intercept-resend attack strategies** The necessity of an authenticated channel between A and B is due to this kind of side-channels. The parties must be able to prove that they are who they say they are, and a man-in-the-middle attack acts exactly at this level: in absence of an authentication channel, E could pose as B and get the secret key directly from A. A similar but more subtle attack is the intercept-resend attack, which consists in E measuring out every signal coming from Alice and then transmitting the measurement results through a lossless classical channel to Bob, ruining the possibility of exploiting entanglement to create a key. A possible solution to this second attack is to move the communication to space, making it more difficult to access the physical medium containing the message.

- **Trojan horse and backflash attacks** A Trojan Horse Attack (THA) consists in E sending large pulses of photons towards one of the setups of the

[4]In quantum cryptography happens often that people create protocols which require considerable technological advances to exist outside of a paper, and then worry about attacks/loopholes that require much more advanced instruments.

two parties and performing measurements on the reflected photons. Indeed, if two parties are using the SARG04 and it's not possible to do PNS attacks, E could still target B shining light onto their device to deduce the basis B will be using, and then implement a intercept-and-resend attack before A's photons arrive, allowing for a complete control over both the bits' values and which bits will be discarded. To protect against this kinds of attacks one can put attenuator between the quantum channel and the device.
Backflash attacks are in principle similar, and they take advantage of a characteristic of avalanches photodiodes (APD), which are used as detectors in these cryptographic protocols. Sometimes APDs emit light, thus the name *backflash*, when detecting a photon, and the polarization of this light can be used to infer the setup of B for example. Introducing spectral filters reduces the backflash light emission.

- **Faked states and detector efficiency attacks** Another assumption which is pivotal to proving the security of protocols such as the BB84 or the BBM92, is the assumption the B and E make independent choices of the measurement basis. This opens up the possibility of side-channels attacks related to imperfections in B's devices which E can exploit to correlate (or control) B's choice of the measurement basis. These schemes use *faked states*, where E instead of gaining information without disturbing the signal aims at giving B the results they want, for example a faked state attack with passive basis selection consists in manipulating the polarization-dependent beam-splitter of B to decide to which basis the photons go.
  Another faked state attack exploits the *detector efficiency mismatch* (DEM), where a detector has different efficiencies for the outcome 0 and 1 of a certain parameter. The attack starts with an intercept-and-resend scheme, where E resends to B a photon in the opposite basis to the one used for the interception, with a timing such that the photon is more likely to give no result rather than give rise to an error; this is called "blinding a detector".

2.1.3 Device-Independent Quantum Key Distribution

Looking at the previous list of side-attacks, it is clear that a big role is due to the devices and how they function: QKD security proofs require certain standards that realistic devices rarely meet. However it's not necessary that it works this way and one can create *device-independent* protocols, which are much less vulnerable, relying mainly on the input-output behavior to certify that the devices work well enough during the protocol.
Since the devices are now black boxes, characterized only by some probability distribution, to prove security one can use Bell nonlocality. In fact if the parties

perform CHSH-like games or violate Bell-like inequalities, that is a certification of randomness in the outcomes: this idea is not new, as shown in Section 2.1.1 the E91 protocol uses exactly this principle. The term device-independent was coined in [1], where the authors created a device-independent protocol secure against collective attacks, i.e. when the only assumption is that E attacks all particles of A and does so always in the same way. A very successful way to have DI-QKD is to use self-testing distributions, a topic explored more in depth in Section 3.1.3.

Up until now I described what is known as *full* device-independence, but this is not the only kind of device-independence that exist, following are the most common types

- *one-sided-device-independence*: the device on one party is completely characterized, while those of the other party are completely uncharacterized.
- *semi-device-independence*: the device is only partly characterized.
- *measurement-device-independent*: A and B (which are both trusted) send quantum signals to a central relay over which no assumption is made, allowing to forego the process of trusting devices and measurements. While the trusted parties have full control over the state which gets sent, E could easily be the central receiver, but it is still possible to extract some secrecy from this setup[5].

2.1.4 Composable security

The last concept regarding QKD which I want to mention is *composability*. It is often the case that a cryptographic task is part of a larger protocol, and it is important to ensure that each task is, not only secure by itself, but it is still secure when it becomes part of a bigger protocol. To assess the composability of a protocol one has to introduce a composable framework, such as the Abstract Cryptography framework, however this goes outside of the scope of this work (for further details [57] contains a clear and formal explanation), so skipping over this we can say that, citing [32] in Section II.D,

> If two protocols are proven secure under a composable security definition the security of their combination can be argued based on their individual functionalities and without the need to give a separate security proof for the combined protocol.

[5] An example of a simple idealized MDI-QKD scheme is in Section IV.G of [32].

An even stronger idea is that of *universally* composable security ([11]), which means that a task is secure by itself and when part of a larger *arbitrary* protocol, i.e. even if some other parts of the protocol are not secure and some information gets leaked to adversaries a universally composable protocol remains secure and secret nonetheless.

2.2 Classical post-processing

Classical post-processing is a way to transform two weakly correlated and possibly only-partially secure keys into an identical couple of highly secure keys, which the parties can then use to exchange information. In this section I will present some of the most common subprotocols which are part of the classical post-porcessing of a QKD protocol, following the much more detailed work [34].

2.2.1 Information Reconciliation

Information reconciliation is a protocol that allows to raise the correlation between classical values which are weakly correlated leaking a minimal amount of information to the eavesdropper. In Section 6.3 of [34] is given a definition more adherent to the protocol of Section 4, where the author considers that A holds a value x and B outputs a guess $\hat{x}$ of x. Calling $\mathcal{X}, \mathcal{Y}$ the alphabets of A and B we have:

Definition 2.2.1 (**IR protocol**)**.** An IR protocol is a two-party protocol where A and B take classical inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively and where B outputs a classical value $\hat{x} \in \mathcal{X}$ or aborts.

The security of an IR protocol is given in terms of a security parameter ε_{IR}, and introducing the space of joint probability distributions $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have:

Definition 2.2.2. Given a probability distribution $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ let $\varepsilon_{\text{IR}} \geq 0$, then an IR protocol is ε_{IR}-*secure on* P_{XY} if the probability that $\hat{x}$ is different from x is less than ε_{IR}, i.e. $\Pr[\hat{x} \neq x] \leq \varepsilon_{\text{IR}}$. Furthermore it is ε_{IR}-*fully secure* if the previous proposition holds for any $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$.

Lastly, the *leakage* will be related to the set of all possible communications[6] executed during the protocol. In other words it is the amount of secret information which correlates to public information, allowing the eavesdroppers to gain knowledge.

[6] A formal definition, unnecessary here, is Definition 6.3.2 in [34].

An important result stemming from information reconciliation, is what is known as Shannon coding theorem, which puts in relation the leakage and the Shannon entropy:

Theorem 2.2.1 (**Shannon coding theorem**). *Let $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a probability distribution and let $n, \varepsilon \geq 0$, then there exists an* IR *protocol which is ε-fully secure and ε-robust on the probability distribution $P_{X^n Y^n} = (P_{XY})^n$ with leakage*

$$\text{leak}_{\text{IR}} \leq nH(X|Y) + \sqrt{3n \log(2/\varepsilon)} \log(|\mathcal{X}| + 3) \tag{2.1}$$

The idea of robustness is related to how sensitive is the protocol to noise, i.e. how much noise the protocol can tolerate before being aborted.

The proof of this theorem requires a longer treatment of the subject and the introduction of a handful of other theorems, lems and specific protocols, but this is nonetheless important since it gives a better idea of the relation between the correlation of two bits and the associated leakage.

The last concept which is useful to have at hands is the *quantum bit-error rate*, or *QBER*. Let us imagine for a second that the channel over which the parties are communicating is a noisy one, i.e. there could be some errors due to dephasing or flipping or other changes in the bits, then it very well may be that the users detect an error even if E is not on the line. The percentage of error that the parties may experience is represented by the QBER, in other words it can be interpreted as the probability that two bits in the raw keys of A and B differ. Information reconciliation (or in this context error correction, which is equivalent) allows to address and correct this kind of issues as well.

2.2.2 Privacy Amplification

The next step, and a fundamental problem of cryptography, is to distill a secret key from the partially secret data that was used to create the *raw* key, and privacy amplification (PA) exists exactly to address this task. A central result of PA was to prove that the key resulting from a hashing of the raw key is secure under only two conditions: the hashing must be done choosing at random a function from a two-universal family of functions, and the length of the distilled key must be shorter than the uncertainty of the adversary on the input in terms of (smooth) min-entropy. This is summarized in the Leftover-Hash Lemma, however before presenting it we must define some mathematical concepts, to make it more digestible.

Definition 2.2.3 (**Two-universal hashing**). Let $\mathcal{F}$ be a family of functions from $\mathcal{X}$ to $\mathcal{Z}$ and let P_F be a probability distribution on $\mathcal{F}$; the pair $(\mathcal{F}, P_F)$ is called *two-universal* if taken any two elements x, x' of $\mathcal{X}$ at random and any function f at random from $\mathcal{F}$ the collision probability is less than the inverse of the dimension of the alphabet $\mathcal{Z}$, i.e. $\Pr\big[|f(x) = f(x')|\big] \leq 1/|\mathcal{Z}|$.

This simply means that the probability that the hashing, which is basically a masking of the message x, makes two different messages unrecognizable, must be lower that the number of possible messages I can get doing the hash.
At this point the leftover-hash lemma states:

> Given two random variables X and Y and taken at random a function f from $\mathcal{F}$, the resulting $Z = f(X)$ will be on average ε_{LHL}-distant from uniform conditioned on Y, with $\varepsilon_{\text{LHL}} = \frac{1}{2}\sqrt{2^{l - H_{min}(X|Y)}}$

which can be put into a formula: for any $\varepsilon > 0$ there is a function f which extracts

$$l \leq \left\lfloor H_{min}^{\varepsilon_{\text{smooth}}}(X|Y) - 2\log\left(\frac{1}{2\varepsilon}\right)\right\rfloor \tag{2.2}$$

bits ε_{LHL}-close to uniform and independent of Y[7].

This is actually a particular case of a more general situation, indeed in [50] we have a quantum version of the LHL. First define a δ-almost two-universal family of hash functions

Definition 2.2.4. Let $\mathcal{F}$ be a family of functions from $\mathcal{X}$ to $\mathcal{Z}$ and let P_F be a probability distribution on $\mathcal{F}$; the pair $(\mathcal{F}, P_F)$ is called *δ-almost two-universal* if taken any two elements x, x' of $\mathcal{X}$ at random and any function f at random from $\mathcal{F}$ the collision probability is less than δ, i.e. $\Pr\big[|f(x) = f(x')|\big] \leq \delta$. Choosing $\delta = 1/|\mathcal{Z}|$ recovers the two-universal function family.

Given this extension the authors define a General Leftover Hash Lemma

Theorem 2.2.2 (**General LHL**). *Let X be a random variable and E be a quantum system, let $\mathcal{F}$ be a δ-almost two-universal family of hash functions from $\mathcal{X}$ to $\{0,1\}^l$, then choosing a random function f from $\mathcal{F}$ the output $Z = f(X)$ is ε'-close to uniform conditioned on E, where*

$$\varepsilon' = \inf_{\varepsilon > 0} \frac{1}{2}\sqrt{(2^l\delta - 1) + 2^{l - H_{min}(X|E) + \log(2/\varepsilon^2 + 1)}} + \varepsilon$$

[7] All the logarithms are base 2.

Furthermore if $\delta < 2^{-l}$, i.e. if $\mathcal{F}$ is two-universal we have

$$\varepsilon' = \frac{1}{2}\sqrt{2^{l-H_{\min}(X|E)}}$$

which recovers the LHL as in (2.2), showing how this definition is more general.

The relevance of the Leftover Hash Lemma is that, in the protocol presented in Section 4 for example, once one calculates the min-entropy and chooses a distance from uniform, they can easily extract the maximal obtainable length of the distilled key.

2.3 Quantum Conference Key Agreement

A generalization of QKD is *Conference Key Agreement* (*CKA*) or *multipartite key distribution*, where more than 2 parties wish to establish secure communication, allowing for the broadcasting of messages in a network of users. Conference key agreement can be achieved in many different ways, for example one possibility is to create bipartite keys between couples of parties and then share a secret common key amongst everybody using the keys already possessed, however this is not efficient in the classical case. Quantum resources allow to use the advantages of QKD, and conference key agreement with such resources is called *Quantum Conference Key Agreement* (*QCKA*). This section follows mainly [28].

To have QCKA we must first establish sources of multipartite entanglement, which are more convoluted[8] than the bipartite entanglement we've seen up until now. It's still possible to qualitatively grasp the idea with some examples: two of the simplest sources of multipartite entanglement are GHZ and W states. For n parties we have:

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}) \tag{2.3}$$

$$|W\rangle = \frac{1}{\sqrt{n}}(|100\ldots0\rangle + |010\ldots0\rangle + \cdots + |00\ldots01\rangle) \tag{2.4}$$

[8] A detailed treatment requires the introduction of concepts such as the *k-separability* and the *genuine multipartite entanglement*, however there is no need for such a treatment here.

where $|i\rangle^{\otimes n} = \underbrace{|i\rangle \otimes \cdots \otimes |i\rangle}_{n \text{ times}}$. For $n = 3$ for example we have

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$$
$$|\text{W}\rangle = \frac{1}{\sqrt{3}}(|100\rangle + |010\rangle + |001\rangle)$$

An important difference is in the robustness to particle loss of these two states, in fact the GHZ state is weak to this kind of noise, losing its entanglement if one of the particles is lost, while the W state still exhibits bipartite entanglement if any one of the particles is lost. This can be easily shown: assume we lose the third particle in both the GHZ and W state, then

$$\begin{aligned}\rho_{\text{GHZ}} &= \frac{1}{\sqrt{2}}\text{Tr}_3\big[\,|000\rangle\langle 000| + |111\rangle\langle 000| + |000\rangle\langle 111| + |111\rangle\langle 111|\,\big] = \\ &= \frac{1}{\sqrt{2}}(|00\rangle\langle 00| + |11\rangle\langle 11|)\end{aligned} \tag{2.5}$$

where I used the linearity of the trace and the orthogonality of the two states, while

$$\begin{aligned}\rho_{\text{W}} = \frac{1}{\sqrt{3}}\text{Tr}_3\big[\,&|100\rangle\langle 100| + |010\rangle\langle 100| + |001\rangle\langle 100| + |100\rangle\langle 010| + \\ &+ |010\rangle\langle 010| + |001\rangle\langle 010| + |100\rangle\langle 001| + |010\rangle\langle 001| + |001\rangle\langle 001|\,\big] = \\ &= \frac{1}{\sqrt{3}}(|10\rangle\langle 10| + |01\rangle\langle 10| + |10\rangle\langle 01| + |01\rangle\langle 01| + |00\rangle\langle 00|)\end{aligned} \tag{2.6}$$

In matrix representation maybe it's easier to see

$$\rho_{\text{GHZ}} = \frac{1}{\sqrt{2}}\begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1\end{bmatrix}\ ,\ \ \rho_{\text{W}} = \frac{1}{\sqrt{3}}\begin{bmatrix}1&0&0&0\\0&1&1&0\\0&1&1&0\\0&0&0&0\end{bmatrix}$$

The reduced density matrix of the GHZ state is diagonal, thus it is just a classical mixture of states, while the one relative to the W state is not diagonal, and diagonalizing it one can show that there is still entanglement.
In [28] are some examples of extensions of QKD protocols, such as the BB84 or the six-state protocol, to multiparty scenarios.

2.3.1 Composable Security

Composable security is implied in the definition for QCKA protocols, and we can show this starting from the definitions of correctness and secrecy. For this purpose consider N parties $A, B_1, \ldots, B_{N-1}$ that want to establish a common secret key.

Definition 2.3.1 (**Correctness**)**.** A CKA protocol is ε_c-correct if

$$p(K_A = K_{B_1} = \cdots = K_{B_{N-1}}) \geq 1 - \varepsilon_c$$

where K_i are the keys held by the parties and $p(K_A = K_{B_1} = \cdots = K_{B_{N-1}})$ is the probability that the keys are identical.

Definition 2.3.2 (**Secrecy**)**.** A CKA protocol is ε_s-secret if

$$\frac{p(\Omega)}{2}||\rho_{K_A E|\Omega} - \tau_{K_A} \otimes \rho_{E|\Omega}|| \leq \varepsilon_s$$

where $p(\Omega)$ is the probability that the protocol does not abort, $\rho_{K_A E|\Omega}$ is the state shared by A and the eavesdropper E once the protocol is done, and τ_{K_A} is the maximally mixed state over all possible values that the key can assume.

The secrecy requirement states that A's key is randomly chosen among the set of possible strings and the eavesdropper has no information about the key except for probability at most ε_s. Once these concepts are defined we can say that a QCKA protocol is ε_{cs}-correct-and-secret if it is ε_c-correct and ε_s-secret for all $\varepsilon_{cs} \geq \varepsilon_c + \varepsilon_s$. We finally can define

Definition 2.3.3 (**Security of a QCKA protocol**)**.** A QCKA protocol is $(\varepsilon_{cs}, \varepsilon_C)$-secure if it is

- *Sound*: for any implementation of the protocol it is ε_{cs}-correct-and-secure.
- *Complete*: there exists an honest implementation of the protocol such that the probability of not aborting is $\geq 1 - \varepsilon_C$.

Definition 2.3.3 implies composable security ([11]).

2.3.2 Device-Independent Conference Key Agreement

Device-independence as a concept was already introduced in Section 2.1.3, and conceptually it is not different: the parties' objective is to certify security using only the observed statistics of inputs and outputs. However differently from the

the DIQKD case, the DICKA scenario is much more complex because an analysis against coherent attacks[9] must comprehend the scenario where E makes the devices behave differently at each round of the protocol. In particular the measurement devices could have quantum memories and could create correlations with the outcomes of previous rounds.
Making use of the Entropy Accumulation Theorem ([15]) it is possible to have DICKA protocols in an adversarial scenario, however it cannot grant the universal composability that is implied when defining non-device-independent CKA, in particular to retrieve universal composability one should not re-use the same measurement devices. A fully DI-CKA protocol was first proposed in [37].

[9] A coherent attack is the most general type of attack, where the only constraint on the eavesdropper is that they respect the laws of quantum mechanics.

Chapter 3

Quantum Networks

In this chapter I will expand on the topics presented in Section 1.3, going beyond the standard Bell scenario. I will refer mainly to [47].

3.1 Overview of quantum networks

A quantum network is a collection of parties which receives information from connected sources, thus they come with an essential difference from the standard Bell scenarios: they have more than one source. This feature allows for an arbitrariness in the configuration of the network (Fig. 3.1) which is not possible in the (2, i, o) scenarios (2 parties, i inputs and o outputs), and furthermore there are results and considerations which are completely new and native of network settings.

The first question that must be addressed is "are the sources independent?" The answer is positive, all sources are independent in a network, and while this is an additional assumption it is a natural one when considering distant sources that are built and operated separately, even if cryptographically is a strong statement. The above question did not make any sense in the standard Bell scenario, while in this context is completely legitimate and has important consequences. For the scenario in Fig. 3.1**(a)** there are two possible interpretations for what the local hidden variable λ could be: it can be a physical property associated with particles which have a common past **or** a free and unlimited source of local randomness available to the parties. This equivalence does not hold anymore for multisource scenarios, as the global randomness is not anymore a free resource. Indeed in a situation like the one in Fig. 3.1**(a)**, assuming a LHV λ exists, the joint probability distribution is

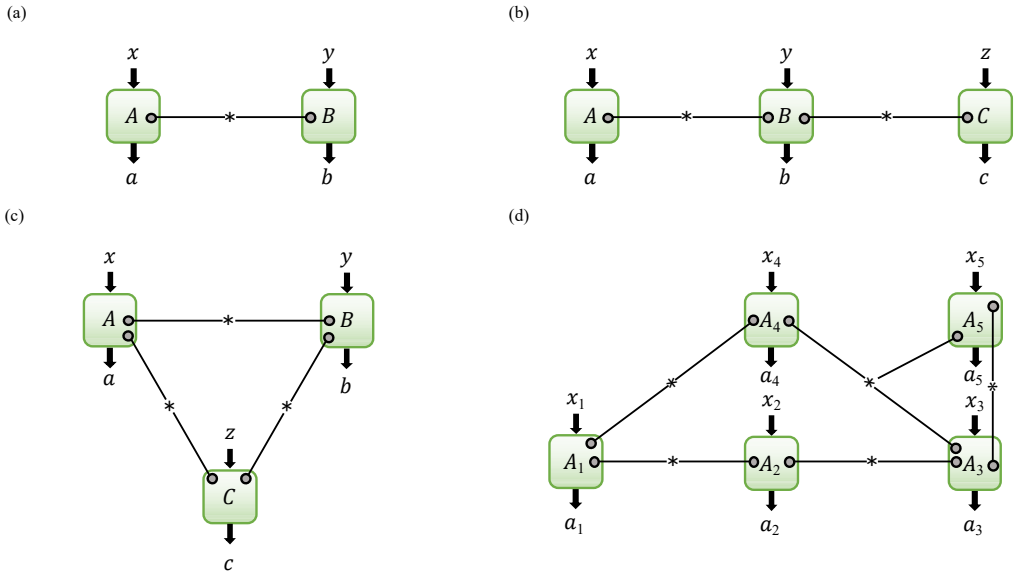

Figure 3.1: **(a):** Standard Bell scenario, 2 parties (A, B) which share a common (bipartite) source, 2 inputs (x, y) and 2 outputs (a, b). **(b):** Bilocal scenario, two parties share a bipartite source each with a central party. **(c):** Triangle scenario, three parties are connected to each other by two bipartite sources each. **(d):** A more complex network, made up of multipartite sources and cycles, which makes some parties (such as A_5 with A_1 or A_2) conditionally independent ([47], p.4).

$$p(a, b|x, y) = \int d\lambda \mu(\lambda) p(a|x, \lambda) p(b|y, \lambda) \tag{3.1}$$

where $\mu(\lambda)$ is the density of the LHV and λ contains all the randomness. This means that if more randomness was introduced to the experiment, for example A and B choose x and y randomly, I can always introduce a new LHV λ' which does not change the overall shape of the local model in (3.1). Now let us look at Fig. 3.1**(b)**, the probability distribution is

$$p(a,b,c|x,y,z) = \int d\lambda_{AB} d\lambda_{BC} \mu_1(\lambda_{AB}) \mu_2(\lambda_{BC}) \cdot \\ \cdot\, p(a|x,\lambda_{AB}) p(b|y,\lambda_{AB},\lambda_{BC}) p(c|z,\lambda_{BC}) \tag{3.2}$$

and it is clear that adding global randomness in the same way as before would correlate the two sources, violating the assumption of independence!
Even if this treatment will be mostly about bipartite sources (connected to 2 parties), to write the most general network local model it's necessary to use multipartite sources: consider $\bar{a}$ the set of all outcomes $(a_1, \ldots, a_n)$, $\bar{x}$ the set of all inputs $(x_1, \ldots, x_n)$[1], $\bar{\lambda}_j$ the set of all LHVs received by the j-th party and $\mu_i(\lambda_i)$ the probability density of the LHV associated to the i-th source. Then the network local model is

$$p(\bar{a}|\bar{x}) = \int d\lambda_1 \mu_1(\lambda_1) \cdots \int d\lambda_m \mu_m(\lambda_m) p(a_1|x_1, \bar{\lambda}_1) \ldots p(a_n|x_n, \bar{\lambda}_n) \tag{3.3}$$

3.1.1 Polytopes for networks

Before going further and studying nonlocality I will briefly present some properties of the of the set of local correlations that can be obtained in a network using the polytope formalism introduced in Section 1.3.3:

1. The network local polytope is contained in the local polytope.
2. The network local polytope is not convex.
3. The network local polytope is connected and characterized by a finite number of polynomial inequalities, i.e. it's closed.

Property 1. is trivial since the local polytope is what one would obtain relaxing the source independence assumption. Property 2. is the most important one: the non-convexity of the set of local correlations implies that the optimal Bell inequalities are no longer the hyperplanes which separate different correlations, and makes the identification of nonlocality a much harder feat when compared to standard Bell scenarios, both analitically and numerically. The previous property implies that network Bell inequalities are non-linear, however Property 3. comes to the rescue, assuring us that there is only a finite number of non-linear Bell

[1]While it is not necessary that the parties have the same number of inputs and outputs, it is common practice to only have one input and one output per party.

inequalities that characterize the local polytope, which can be restricted to polynomial expressions ([18]). An example of network polytopes is in Fig. 3.2**(a)**. In the quantum model the independence of sources translates into the condition of having one state per source, and while the state can be entangled itself, there can be no entanglement or classical correlation between states of different sources. Similarly to the classical case some properties are:

4. The quantum network set is contained in the quantum set and contains the network local polytope.

5. The quantum network set is non-convex.

6. The quantum network set is closed.

and are graphically represented in Fig. 3.2**(b)**.

Finally, there exist correlations that can be generated with quantum resources which is not possible to reproduce with the same number of parties sharing a global local variable and vice versa, meaning that the quantum network set does not contain the local polytope and vice versa.

3.1.2 Genuine Network Nonlocality (GNNL)

Until now I did not give a definition of network nonlocality, and I have not explained how this differs from the usual Bell nonlocality, pivotal to the standard quantum mechanics courses in any physics faculty. Unfortunately it is not as easy as saying "there exist no network local model compatible with the set of correlations", and this is a consequence of the fact that genuine network nonlocality manifests itself in multiple different ways, some of which can be traced back to standard Bell nonlocality. The next example shows such a scenario: imagine that three parties are in the bilocal configuration of Fig. 3.1**(b)**, and in this scenario we have A and B doing CHSH tests on their marginal distribution, while C outputs random numbers each round. The local model for the marginal of A and B is

$$\begin{aligned} p(a,b|x,y) &= \sum_c \int d\lambda_1 d\lambda_2 \mu_1(\lambda_1)\mu_2(\lambda_2) p(a|x,\lambda_1) p(b|y,\lambda_1,\lambda_2) p(c|z,\lambda_2) = \\ &= \int d\lambda_1 \mu_1(\lambda_1) p(a|x,\lambda_1) \int d\lambda_2 \mu_2(\lambda_2) p(b|y,\lambda_1,\lambda_2) = \\ &= \int d\lambda_1 \mu_1(\lambda_1) p(a|x,\lambda_1) p(b|y,\lambda_1) \end{aligned} \tag{3.4}$$

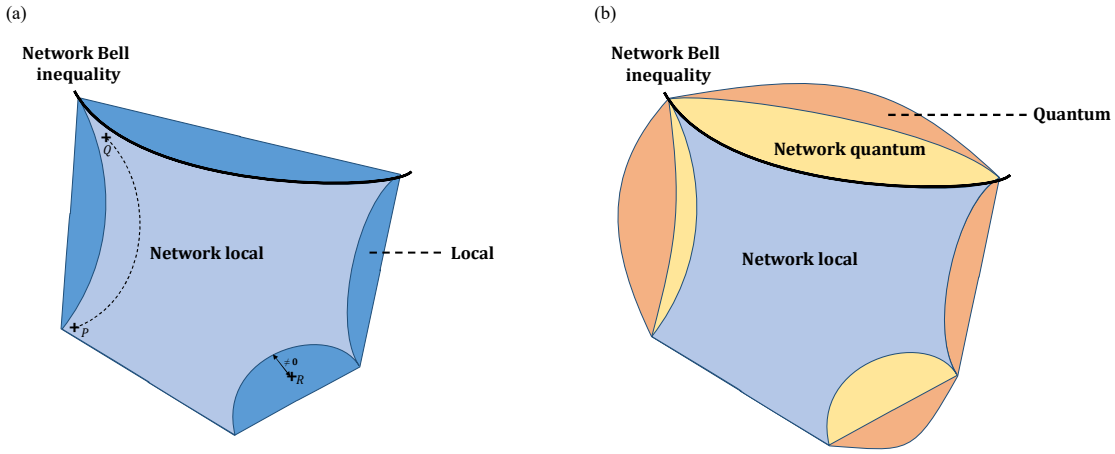

Figure 3.2: **(a):** The relations between local and network local polytopes, i.e. there exist classical correlations that violate network local correlations, which in turn means that when considering only local resources a violation of network Bell inequalities is a certification that the network has a different structure than the one hypotesized. **(b):** Relations between the set of quantum correlations, network quantum correlations and network local polytope. Once the network structure is fixed network Bell inequalities acquire a similar meaning to Bell inequalities. ([47], p.5)

which is unsurprisingly the standard local model. This means that, in presence of a violation of the CHSH tests performed by A and B, an external observer examining the distribution $p(a, b, c|x, y, z)$ could mistake it to possess genuine network nonlocality (GNNL), while it is only (masked) Bell nonlocality. This was just a simple example, but the same idea applies to any network, thus the need to have a different formulation of GNNL. A definition was first given in [44] in an operational way, but, in order to avoid complex treatments and concepts, here I will only give a basic definition in line with the previous example: considering Bell nonlocality as a resource, correlations that can provably not be obtained by these resources are termed genuine network nonlocal.

In simple cases such as the bilocal or the triangle scenarios(Fig. 3.1**(b)** and **(c)** respectively), showing the presence of genuine network nonlocality is fairly simple

and there are many ways. Following is one method to prove nonlocality in the first scenario, while in Section 3.2 I will treat more in depth the second case.
In [47] Section 2B are three ways to show the presence of GNNL in the bilocal scenario, and here I will follow the *Elegant Joint Measurement* (EJM) approach, rather than the more common violation of Bell inequalities using Bell states or the explicit bilocal decomposition. The EJM is a kind of entanglement swapping measurement introduced in [20] to encode information in antiparallel spins, but I will follow the notation and formalism of [47] instead of the original paper. The EJM is a projection on the basis

$$|e_1\rangle = \frac{1}{\sqrt{8}}(-1-i, 0, -2i, -1+i)^\dagger \ , \quad |e_2\rangle = \frac{1}{\sqrt{8}}(1-i, 2i, 0, 1+i)^\dagger \tag{3.5}$$
$$|e_3\rangle = \frac{1}{\sqrt{8}}(-1+i, 2i, 0, -1-i)^\dagger \ , \quad |e_4\rangle = \frac{1}{\sqrt{8}}(1+i, 0, -2i, 1-i)^\dagger$$

which is orthonormal and complete. Measuring in this basis allows for quantum correlations that violate the bilocal Bell inequality introduced in [46], where the authors setup a system as in Fig. 3.3

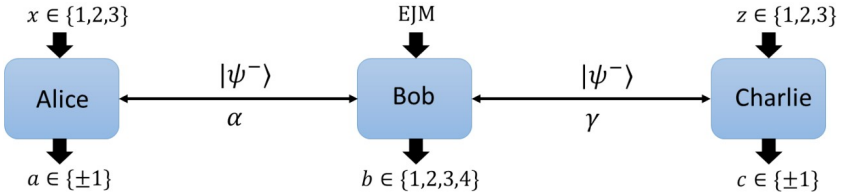

Figure 3.3: Setup for the violation of the TGB (Tavakoli-Gisin-Branciard) Bell inequality. Both A and C have one input with three possible values and binary outcomes, while B always measures using the EJM and has quaternary outcome. The sources distribute the maximally entangled state $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. ([46], p.1)

In this context the probability distributions can be expressed in terms of (single-, two- and three-party) correlators[2], which can then be combined into the bilocal Bell inequality

$$\mathcal{S}_{TGB} = \frac{1}{3}\left(\sum_{y=z}\langle B^y C_z\rangle - \sum_{x=y}\langle A_x B^y\rangle\right) - \sum_{x\neq y\neq z\neq x}\langle A_x B^y C_z\rangle \leq 3 \tag{3.6}$$

[2]The explicit form of the correlators is not relevant for this qualitative treatment, the explicit derivation of the bilocal Bell inequality is in [46].

Using the Elegant Joint Measurement in this setup one obtains $\mathcal{S}_{TGB} = 4$, which violates the previous inequality, while no known protocol based on Bell state measurement achieves a violation of (3.6). Furthermore, even if an analytical proof is still lacking, numerical proofs support the hypotesis that this holds for general bilocal models as well.

3.1.3 Self-Testing and Token-Counting

The last topic of this section is self-testing. Self-testing is, in simple terms, a method to infer the underlying physics of a quantum experiment in a black box scenario, as stated in [45]. In this terms nothing alludes to quantum networks, simply because self-testing is not inherent to quantum networks, however after a short formal introduction, the treatment will be only relative to self-testing in network scenarios. For this section I will mainly follow the review [45].

Self-testing is the most powerful method to certify device-independence in quantum information theory, consequence of a particular feature of nonlocality: some nonlocal correlations can be produced uniquely by making specific sets of incompatible measurements on specific entangled states. This allows to certify certain quantities in a device-independent way, such as the underlying entangled state or the measurements up to local isometries, all through Bell inequalities and nonlocality.

Definition 3.1.1 (**Self-testing of states and measurements**)**.** The correlations $p(a, b|x, y)$ self-test the state $|\psi'\rangle_{A'B'}$ and measurements $\{M'_{a|x}\}, \{N'_{b|y}\}$ if for any state ρ_{AB} and measurements $\{M_{a|x}\}$, $\{N_{b|y}\}$ compatible with $p(a, b|x, y)$ and for any purification $|\psi\rangle_{ABP}$ of ρ_{AB}, exists a local isometry $\Phi_A \otimes \Phi_B$ such that

$$\Phi_a \otimes \Phi_B \otimes \mathbb{1}_P \left[M_{a|x} \otimes N_{b|y} \otimes \mathbb{1}_P |\psi\rangle_{ABP}\right] = \left(M'_{a|x} \otimes N'_{b|y} \otimes \mathbb{1}_P |\psi'\rangle_{A'B'}\right) \otimes |\xi\rangle_{\bar{A}\bar{B}P}$$

for all a, b, x, y and some state $|\xi\rangle_{\bar{A}\bar{B}P}$.

Measurement operators and the purification of a state were previously introduced, (Section 1.2 and 1.1.1 respectively), $|\xi\rangle_{\bar{A}\bar{B}P}$ is a junk state so we are not really interested in defining it further, and Φ are isometries, meaning maps that preserve distances.

A specific kind of strategy which leads to a family of self-testing distributions (of which an example is treated in Section 3.2.1), are the Token Counting (TC) strategies: these are characterized by the fact that a source randomly emits a fixed number of tokens, of which i are sent right with probability $p(i)$ and the others are sent left, successively the parties output the total number of tokens they received,

and the resulting correlations are called Token Counting distributions. Following [35] I will present some concepts which will be central in the next section, starting with a formal definition of a TC strategy.

Definition 3.1.2 (**TC strategy**)**.** A strategy $\mathcal{S}$ is called Token-Counting if the following statements hold up to local isometries (relabeling of the outputs and information sent by the sources):

1. Each source S_i distributes a fixed number of tokens N_i to the connected parties, following some fixed probability distribution and possibly along with additional information.

2. Each party A_j counts and outputs $a_j = (n_j, \alpha_j)$ where n_j is the total number of received tokens and α_j is possibly additional information.

The resulting distribution $P[(n_1, \alpha_1), \ldots, (n_J, \alpha_j)]$ is a Token-Counting distribution.

In Theorem 1 of [35] is the proof that for some networks, TC strategies are the only classical strategies that produce TC distributions, with the constraint that the network must belong in the class of *No Double Common-Source* (NDCS) networks

Definition 3.1.3 (**No Double Common-Source networks**)**.** A network $\mathcal{N}$ is NDCS if taken each couple of parties they share at most one source.

Looking at Fig. 3.1 for example the first three cases are obviously NDCS, while in **(d)** A_3 and A_5 share two sources, so for that network TC distributions could be achieved with non-TC strategies.
Finally we can define *rigidity* as a consequence of Theorem 1 of [35].

Theorem 3.1.1. *Let $\mathcal{N}$ be a NDCS network, then any classical strategy that simulates a TC distribution in $\mathcal{N}$ is necessarily a TC strategy. Moreover the probability distribution of the tokens emitted by the sources in any such TC strategy is unique.*

The proof is in Appendix A of [35]. Based on this we have an informal definition of rigidity: fixed a network $\mathcal{N}$ a TC distribution is **rigid** if it can only be reproduced using a specific TC strategy.
This result can be generalized to quantum networks (formalized in Result 2 and proven in Appendix C of [41]) and we have quantum TC strategies, where the sources distribute superpositions of states and the parties perform projective measurements in order to understand how many tokens they received. These quantum TC strategies give rise to TC distributions which cannot be simulated classically.

3.2 Triangle Network

The simplest network without conditional independence is the triangle network, shown in Fig. 3.1**(c)**, where three parties are connected pairwise by independent bipartite sources. The network local model in this case is

$$p(a,b,c|x,y,z) = \int d\lambda_1 \mu_1(\lambda_1) \int d\lambda_2 \mu_2(\lambda_2) \int d\lambda_3 \mu_3 \lambda_3 \cdot \\ \cdot\, p(a|x,\lambda_1,\lambda_2) p(b|y,\lambda_2,\lambda_3) p(c|z,\lambda_3,\lambda_1) \tag{3.7}$$

The absence of conditional independence between the parties means that the techniques used for the bilocal scenario, such as the Elegant Joint Measurement, are incompatible and one must prove the presence of nonlocality in some other way.

3.2.1 The RGB4 Distribution

One striking feature of quantum networks is that it makes possible to have nonlocality without inputs. In standard Bell scenarios the inputs introduce the randomness which results in the non-correlation of the outputs, while without inputs the parties would always perform the same fixed measurement, and thus they would observe some correlated outcomes. This is not the case for quantum networks, where the parties can perform always the same fixed measurement and obtain uncorrelated results, thanks to the assumption of independence between the sources, which in this sense "substitutes" the presence of inputs.

A distribution which achieves this feat is the one introduced in [36], known as RGB4 distribution, and it was the first example of a set of correlation that cannot be traced back to Bell-like nonlocality. In Fig. 3.4 is the network under analysis. In the original paper the authors create a slightly different, but completely equivalent, setup to the one used in Chapter 4: A, B, C are the three parties with outputs a, b, c and every pair of observers is connected through a bipartite source sharing a physical system, whether it is a classical variable or a quantum state. The sources are independent and this assumption is, as explained before, crucial. As already mentioned the set of distributions is non-convex, which makes it very difficult to characterize $p(a,b,c)$. The authors decided to change approach and construct a family of quantum distributions, using entangled states and entangled measurements, and then show that for this family there exist no local model with independent sources, i.e. that $\nexists\, \alpha, \beta, \gamma$ such that

$$p(a,b,c) = \int d\alpha \int d\beta \int d\gamma p(a|\beta,\gamma)p(b|\alpha,\gamma)p(c|\alpha,\beta) \tag{3.8}$$

The sources in this scenario all distribute the state $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and each party performs a measurement in the basis

$$\begin{aligned} |\uparrow\rangle &= |01\rangle \ , \ \ |\chi_0\rangle = u\,|00\rangle + v\,|11\rangle \\ |\downarrow\rangle &= |10\rangle \ , \ \ |\chi_1\rangle = v\,|00\rangle - u\,|11\rangle \end{aligned} \tag{3.9}$$

with u, v[3]$\in \mathbb{R}$ and related by $u^2 + v^2 = 1$ and $0 < v < u < 1$.

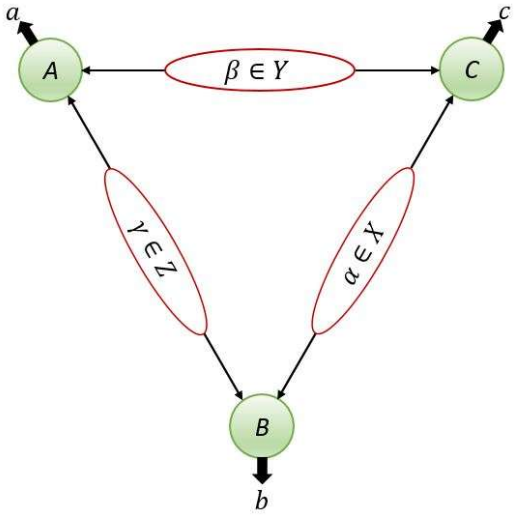

Figure 3.4: Triangle configuration without inputs. In the red ovals are the sources, each associated to a local hidden variable ($\alpha \in X, \beta \in Y, \gamma \in Z$), and in the green circles are the three parties with their outcomes. ([36], p.1)

The proof that this instance of the RGB4 distribution cannot be reproduced by any classical local model (in the sense of (3.8)), for $u^2_{min} < u^2 < 1$ and $u^2_{min} \approx 0.785$, can be found in Appendix A of [36]. However the proof that to have this specific distribution one must have entangled measurements and/or states was achieved in [41], where the authors prove the (quantum) rigidity of RGB4, and find lower bounds on the entanglement of the distributed states, and on the min-entropy

[3]To understand better what these parameters are one must consider an experimental setup where pairs of entangled photons are emitted from the sources and pass through a beam-splitter to reach the two parties. In this context u and v are related to the beam-splitters' transmittivity.

of the local outcome, certifying an intrinsic level of randomness of this distribution.

The setup is:

- The distributed state is $|\psi^+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$.
- The measurement is done in the basis

$$\begin{aligned} &\Pi^0 = |00\rangle\langle 00| \ , \ \ \Pi^{1_0} = |1_0\rangle\langle 1_0| \\ &\Pi^2 = |11\rangle\langle 11| \ , \ \ \Pi^{1_1} = |1_1\rangle\langle 1_1| \\ &\text{with } \ |1_i\rangle = u_i\,|01\rangle + v_i\,|10\rangle \end{aligned} \tag{3.10}$$

  where $u_0 = -v_1 = \cos\theta$ and $v_0 = u_1 = \sin\theta$ for a parameter $\theta \in [0, \pi/4]$.
- Thus the set of outcomes is $\{0, 1_0, 1_1, 2\}$.

This gives rise to the distributions, from now on referred to as RGB4, given by

$$\begin{aligned} &p(1_i, 1_j, 1_k) = \frac{1}{8}(u_i u_j u_k + v_i v_j v_k)^2 \\ &p(1_i, 0, 2) = \frac{1}{8}u_i^2 \ , \ \ p(1_i, 2, 0) = \frac{1}{8}v_i^2 \quad \circlearrowleft \end{aligned} \tag{3.11}$$

where $\circlearrowleft$ means "up to cyclic permutations", and all the other probabilities are exactly zero. A nice feature is that coarse graining the outputs 1_0 and 1_1 into a single outcome 1 makes the distribution Token-Counting[4], which allows for the use of self-testing results, i.e. a device-independent characterization! To sum up in [41] the authors: first show that quantum nonlocal distributions on ring networks without inputs can be (partially) self-tested, and then apply this result to the triangle network, proving that the RGB4 distribution has genuine network nonlocality, an intrinsic randomness in their outcomes and a certain amount of entanglement in the distributed states. These results were central for the development of the cryptographic protocol in Chapter 4.

3.2.2 Inflation

Inflation is a general technique to spot nonlocality in networks, setting constraints on the correlations arising, and while this is not the focus of this work, it still is interesting to give a brief overview and an application to the triangle network.

[4]The parties output 0 if they receive 0 tokens, 1 if they receive one token, whether it is from the right or the left, 2 if they receive a token from each side.

Inflation was initially formalized in [56][5], and in layman's terms consists in determining whether a given probability distribution can be generated over a determined causal structure: in the context of Bell nonlocality in networks this is equivalent to the determining if a probability distribution has a network local model for such a configuration. Inflating a network is a process which follows these steps:

1. Assume the distribution P_{or} can be generated in the network, i.e. assume that there exist a set of random variables and deterministic functions of such variables that reproduce the target distribution.

2. Add copies of the parties and the sources to the network in such a way that, in the inflated network, each copy of the sources connects only to copies of the parties that were connected to the original sources in the original network. See the left image in Fig. 3.5 for clarity.

3. Analyze the inflated network and derive constraints on the probability distributions.

Expanding on step 2., this process allows to match marginals like $P_{inf}(a^{i,j}, b^{j,k}, c^{k,i})$ to the original distribution $P_{or}(a, b, c)$. The additional marginals, those that do not have equivalents in the original network, such as $P_{inf}(a^{1,1}, b^{1,1}, c^{1,2})$[6] in Fig. 3.5 left, must be matched with all the other marginals of the same kind. In our example $P_{inf}(a^{1,1}, b^{1,1}, c^{1,2})$ must coincide to all the marginals like $P_{inf}(a^{i,j}, b^{j,k}, c^{k,l\neq i})$. Finally once one has all the constraints can move on to step 3. and reformulate them as linear programs (which can always be efficiently evaluated) and determine if a P_{inf} respecting these rules exists.

An attentive eye could have already spotted that the generalization to quantum inflation is not easy, since what was given for granted in Fig. 3.5 is that the information can be copied and redistributed amongst multiple parties, an operation which is notoriously impossible in quantum mechanics as stated by the famous *no-cloning* theorem[7]. The solution was found into copying the sources themselves instead of the subsystems they distribute, and then the parties will decide on which set of copies they'll perform the measurement, as shown in the

[5]Even if the formal introduction was done in [56] the first ideas were developed in [22].

[6]This marginal does not reduce to any marginal in the original network because the sources sending information to $A^{1,1}$ and $C^{1,2}$ are two different sources, β^1 and β^2.

[7]The history of the no-cloning theorem is rather convoluted: in 1982 W. K. Wootters and W. H. Zurech as well as D. Dicks separately published the proof, but one and a half year before that G. Ghirardi had already proven it without publishing it. However in 1970 J. L. Park in [31] gave the first proof of a no-go theorem which would later be recognized as the no-cloning theorem.

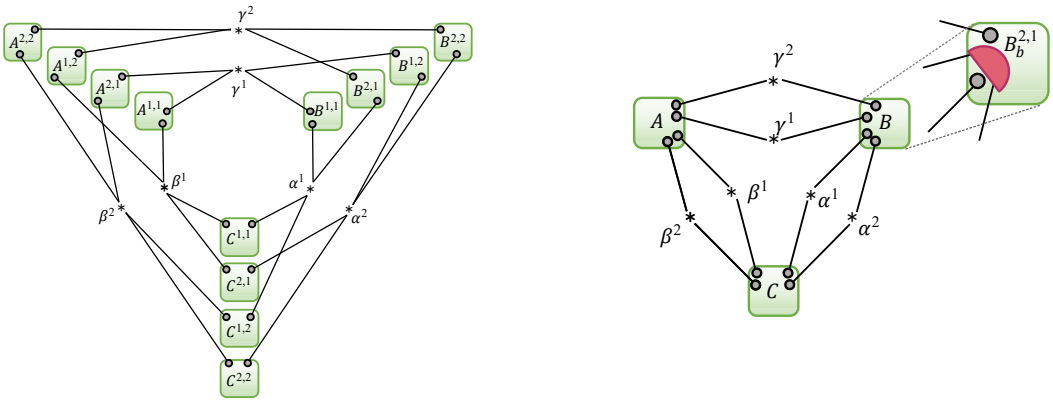

Figure 3.5: Inflation of the triangle network in Fig. 3.1**(a)**. The parties are identified as $X^{i,j}$ where i, j point to the sources from which the party receives states. **Left**: Web inflation of the network, where the arrows for inputs and outputs were removed to avoid a cluttered image. Notice that the same source is connected to multiple parties, but only to those with which it was connected in the original network. ([47], p.20) **Right**: Quantum inflation of the network, in this case the sources cannot copy information, differently from the left side where each source sends the same subsystem to multiple parties, thus avoiding cloning and allowing for identification of nonlocal correlations. ([47], p.21)

right image of Fig. 3.5. As before the problem of finding a (global) quantum state and measurements which correctly reproduce the distribution can be solved using numerical methods, for example via non-commuting polynomial optimization using hierarchies of semidefinite programs, and the non-existence of a solution is a proof of the incompatibility between the distribution and the original quantum network.

Of course there are infinitely many ways to inflate a network, and since they're all equivalent, it is enough to find one for which the distribution does not respect to constraints to prove that there is no network local model. Furthermore in [29] it was proven that there exists a hierarchy of tests which, if are all passed, can guarantee the existence of a network local model.

A short example may be of aid to understand. Look at Fig. 3.6, there are the usual three parties, this time without inputs, and three bipartite sources, and then

the network gets inflated as in the right side, adding a copy of the source γ. The objective is to determine if the following distribution is compatible:

$$P_{GHZ} = \frac{1}{2}(\delta_{A,0}\delta_{B,0}\delta_{C,0} + \delta_{A,1}\delta_{B,1}\delta_{C,1}) = \begin{cases} 1/2 & \text{if } a = b = c \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

where $\delta_{X,x}$ indicates that the device X is related to the outcome x.

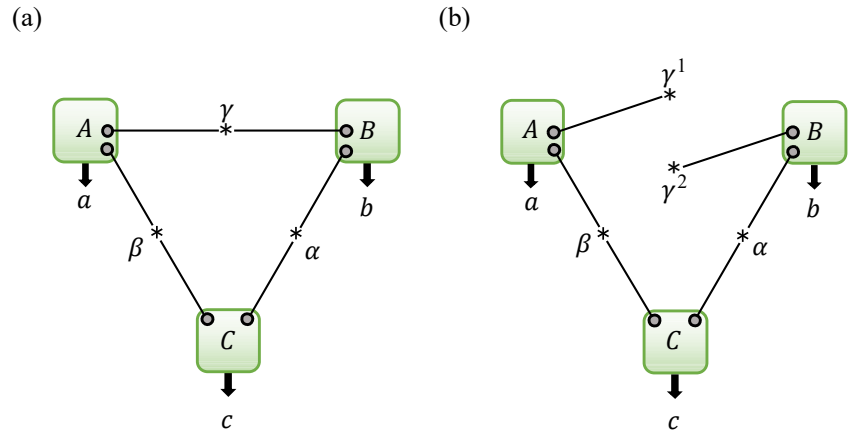

Figure 3.6: **(a)**: Inputless triangle network. **(b)**: The cut inflation of the triangle network, where the source γ is copied and connected only to A or B, creating conditional independence between the two parties. ([47], p.13)

Assume that exist α, β, γ and measurements A, B, C such that they can generate the target distribution P_{GHZ} in a network as in Fig. 3.6**(a)**, and then consider the inflation in Fig. 3.6**(b)**. Analyzing the subnetworks respectively of A and C and then of B and C we have that

$$P^{AC}_{inf} = P^{AC}_{GHZ} = \frac{1}{2}(\delta_{A,0}\delta_{C,0} + \delta_{A,1}\delta_{C,1}) \tag{3.13}$$

$$P^{BC}_{inf} = P^{BC}_{GHZ} = \frac{1}{2}(\delta_{B,0}\delta_{C,0} + \delta_{B,1}\delta_{C,1}) \tag{3.14}$$

while studying the subnetwork of A and B we can notice

$$P^{AB}_{inf} = P^{A}_{inf}P^{B}_{inf} = P^{A}_{GHZ}P^{B}_{GHZ} = P^{AB}_{GHZ} \tag{3.15}$$

since there is no common source between the two parties. These three assumptions are already incompatible: if it never happens that $a \neq c$ nor $b \neq c$ from (3.13), then it must always be $a = b$ since the outcomes are binary, which means that there is a correlation between the outputs of A and B, in contrast to the conditional independence of (3.15), meaning that the assumption that P_{GHZ} have a network local model in the inputless triangle network is wrong.

Chapter 4

Cryptographic Protocol

In this section I will present the original results of the project, carried out with some researchers of the Quantum Information Theory group at ICFO. What follows is a device-independent quantum cryptographic protocol native to quantum networks which does not make use of local randomness.
This last chapter is divided as such: firstly I will show how the protocol can be used to create a key between $M = 2$ parties, A and B, (Section 4.1), and then how it can be used as a quantum conference key agreement (QCKA) protocol to let $M = N$ parties share the same key (Section 4.2). Both of these are done under the assumption of an honest network, i.e. attacks come only from outside the network. Between the two sections is Fig. 4.2, summarizing the QKD protocol in a $N = 4$ network. In Section 4.3 I will prove the security of this protocol, and in Section 4.4 I will present the case where the network is mostly untrusted, i.e. the attacks can come from parties in the network as well. Lastly in Section 4.5 are some open questions and remarks on the protocol.

4.1 Protocol for A and B

4.1.1 Setup

Let the network be a ring of N parties and N bipartite sources, such that each party is connected to a pair of sources in the ring, as shown in Fig. 4.1.

Each source distributes the state $|\psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ and the parties perform the measurement $\{\Pi^0, \Pi^L, \Pi^R, \Pi^2\}$. As stated in Section 3.2.1 we have

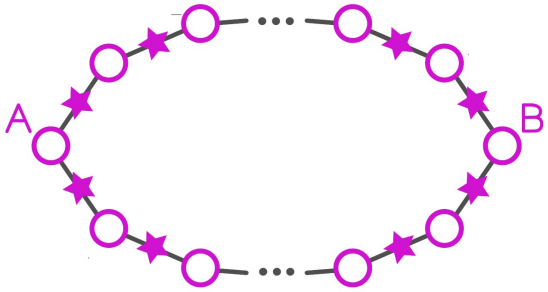

Figure 4.1: The parties are represented by the circles and the sources are represented by the stars.

$$\Pi^0 = |00\rangle\langle 00| \ , \ \Pi^L = |1_L\rangle\langle 1_L|$$
$$\Pi^R = |1_R\rangle\langle 1_R| \ , \ \Pi^2 = |11\rangle\langle 11| \tag{4.1}$$

where

$$|1_R\rangle = u\,|01\rangle + v\,|10\rangle \ , \ |1_L\rangle = v\,|01\rangle - u\,|10\rangle \tag{4.2}$$

with $u \approx \sqrt{0.8}$ and $v = \sqrt{1-u^2}$ both real numbers; the set of outcomes is $\{0, 1_L, 1_R, 2\}$. To make the RGB4 distribution of Section 3.2.1 a Token-Counting distribution, one has to do the coarse graining of the outcomes obeying the map $\{0, 2\} \to 0$, $\{1_L, 1_R\} \to 1$. Having N sources each emitting 1 token, the constraint on the fixed number of total tokens (N) becomes a condition on the parity of the outcomes: namely after the coarse graining there is an even number of outcomes with value 0.

4.1.2 Verification

The first step of the protocol requires the generation of $n = n_{rand} + n_{test} + n_{key}$ rounds, which will be respectively used to create public randomness, verify that there is no tampering going on by evil external parties (E), and to generate a raw key.
Once all the rounds are done and the parties do the measurements, and before the coarse graining happens, B will announce publicly the outcomes of their n_{rand} rounds. This serves two purposes: A can check that the outcome distribution matches the RGB4 (in the case of an infinite number of rounds)[1] and thus decide

[1] Of course in the case n_{rand} finite the outcome distribution will be ε_{LHL}-close to the RGB4 distribution.

whether the protocol can continue or should be aborted. The other purpose is to use the intrinsic randomness of the outcomes, shown to be present in [41], to create a string of random bits S_{test}, using the left-over hash lemma to amplify the randomness if needed, to use to randomly select which of the remaining rounds will be used for eavesdropper detection (n_{test}), and which for the key generation (n_{key}).
At this point A announces their outcomes for the n_{test} rounds and signals to abort the protocol if these deviate from the expected distribution, thereby detecting potential eavesdropping or device imperfections.
These two verification processes allow the parties involved in the key generation to spot eventual attacks from E in any part of the protocol.

4.1.3 Key generation

The next step is an information reconciliation protocol, broken in two elementary steps:

- After the verification part every party does the coarse graining of their own outcomes.
- Every party except from A and B announces their coarse grained outcome.

Using the parity of the sum of these announced outcomes, A (and respectively B) can infer the outcome of B (and respectively A). Based on this inferred outcome, B decides whether to flip their bit or leave it unchanged to establish the shared key. This is possible due to how the set of outcomes is, in fact after the coarse graining the possible outcomes are all the bitstrings of N bits with an even number of 0s. This can be formulated in a more compact way introducing the *Hamming weight* (Δ) of a bitstring, which is simply the number of 1s present, so we have that the possible outcomes are all the permutations of

$$\bar{z} \in \{0,1\}^N : \oplus_i \bar{z}_i \equiv \Delta = N, N-2, \ldots, 0 \tag{4.3}$$

where $\bar{z}$ is the conventional way to represent a string of bits and $\bar{z}_i$ are the elements of said string. It's possible to reach $\Delta = 0$ if and only if N is even, otherwise the smallest possible Δ is 1.
For example in the case of $N = 4$ the set of outcomes is the permutations of $0000, 0011, 0101, 1111$, or

It's clear that knowing only two of the four bits, a malicious party still has to guess between two possible strings of outcomes, while knowing three of the four outcomes gives a unique bitstring. This holds for strings of length N as well: knowing $N-m$ bits the probability of guessing the correct string is $2^{-(m-1)}$, while

A	0	0	0	0	1	1	1	1
B	1	1	0	0	0	0	1	1
C	1	0	1	0	1	0	0	1
D	0	1	1	0	0	1	0	1

Table 4.1: Each column is one possible set of outcomes for four parties. As expected the set of outcomes has $2^N/2 = 2^{N-1}$ elements because of the constraint on Δ.

knowing $N-1$ bits gives full knowledge.

At this point, after every other party in the network has announced their (coarse grained) outcome, A and B know each other's outcome, since $N-2$ bits are public and both of them hold another bit, while anybody but them has to guess between two possibilities. Now both parties can take (for example) A's secret bit a_0 to use as the first bit of the raw key. Repeating the verification and the key generation $\mathcal{S}$ times gives a secret key of length $\mathcal{S}$.

4.1.4 Raw key privacy amplification

The last step consists of increasing the secrecy of the raw key, or equivalently lowering the conditional (von Neumann) entropy $H(K|E)$, where K is the key of length $\mathcal{S}$. During each iteration of the first two steps, resulting in one secret bit, E knows the parity ($\equiv$ bitwise sum $\equiv$ mod2 sum $\equiv$ XOR) between the two secret bits of A and B, due to the constraint on the number of 0s in each outcome. In other words E knows that $a_i = b_i$, where a_i and b_i are A's and B's bits at the i-th step, if $\Delta_i - a_i - b_i = 2n$, and $a_i = b_i \oplus 1$ if $\Delta_i - a_i - b_i = 2n+1$, where $n \in \mathbb{N}$.
The aim is to take the raw outputs O of the parties, over which E has partial knowledge ($H(A|E) < 1$), and do some privacy amplification in order to find a new set of outputs O' which has a smaller correlation (ideally no correlation at all) with E. The cost is that the key will be shorter, i.e. $|O'| < |O|$ where $|\cdot|$ is the cardinality (or length) of the string.
A common way to extract randomness is based on the *leftover hash lemma* (LHL), seen explicitly in Section 2.2.2, recalling the formula

$$l \leq \left\lfloor H_{min}^{\varepsilon_{\text{smooth}}}(X|Y) - 2\log\left(\frac{1}{2\varepsilon_{\text{PA}}}\right)\right\rfloor \tag{4.4}$$

where the logarithms are all base 2. For clarity from now on $\varepsilon_{\text{smooth}} = \varepsilon$ and $\varepsilon_{\text{PA}} = \varepsilon'$. The hardest part is extracting $H_{min}^{\varepsilon}(X|Y)$, because in this case these are not just random variables, but collections (or more appropriately *registers*) of random variables: X is the raw key K and Y is the collection of classical-quantum

information "stolen" by E. The explicit derivation and proof of security is in Section 4.3.2, however following is a qualitative summary of the process leading to the final result:

1. Using the QAEP (1.117) we can lower bound the smooth min-entropy over multiple rounds of the LHL using the von Neumann entropy.

2. From (1.86) and the monotonically decreasing behavior of the quantum Rényi α-entropy we have

$$H_{min}(A|B)_\rho \leq H(A|B)_\rho \leq H_{max}(A|B)_\rho \tag{4.5}$$

   and thus the QAEP becomes

$$H^\varepsilon_{min}(K|\bar{E}) \geq \mathcal{S} \cdot H_{min}(A|E) - \sqrt{\mathcal{S}} \cdot \delta(\varepsilon, \eta) \tag{4.6}$$

   where $\bar{E}$ represent the register of information of E and $\mathcal{S}$ is the length of K. $H_{min}(A|E)$ is the information that E has over a single outcome of A, i.e. a single bit of the raw key.

3. Finally calculating p_{guess} and using the relation in (1.94) yields the last bound required.

Tweaking the values of ε and ε' the parties have control over the secrecy of the key, at the cost of a shorter key.

4.2 QCKA extension

The extension of this protocol to a quantum conference key agreement protocol is slightly different, especially in the key generation part, but maintains the same structure overall.

4.2.1 Setup and verification

The setup is exactly the same as in Section 4.1.1 and shown in Fig. 4.1, but in this case we will call A_0 what before was A and A_i with $i = 1, \ldots, N-1$ the other parties. For the verification part now the process is the same as in Section 4.1.2, there are $n_{rand} + n_{test} + n_{key}$ rounds and measurements, but now all the parties announce publicly their outcomes of the n_{rand} rounds, allowing for the verification of the RGB4 distribution. Then, just as before, A_0 can distill a bitstring to use as seed randomness to verify that no eavesdropping happens in $n_{test} + n_{key}$.

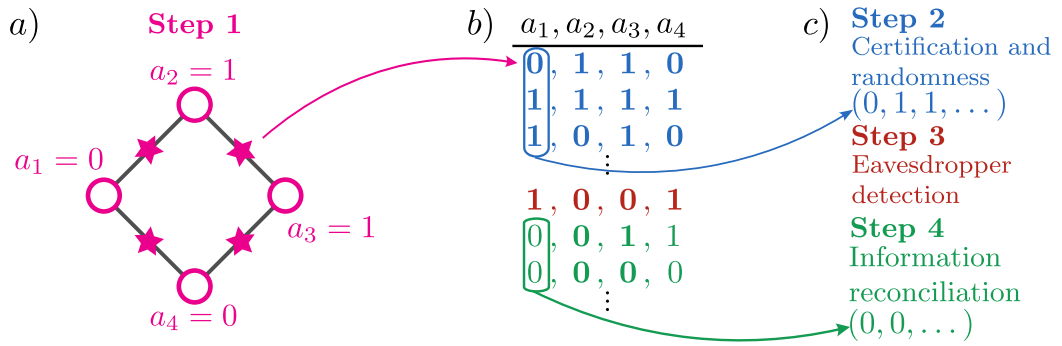

Figure 4.2: Overview of the protocol, in this case $A \equiv A_1$ and $B \equiv A_3$. (a): The setup network, the parties are represented as circles and the sources as stars, with possible outcomes. (b): Over n rounds each party collects outcomes, of which some get announced publicly (in **bold**). (c): Step 2 consists in the verification that the correlations follow the RGB4 distribution to the expected degree, and the extraction of randomness from A_1's outcomes for the eavesdropper detection. Step 3 is when the selection of the eavesdropper detection happens. Step 4 is the key-generation, A_1 and A_3 derive the shared key.

4.2.2 Key generation and privacy amplification

The idea now is to exploit the same process as in Section 4.1.3, but repeated pairwise by A_0 and all the other parties. In addition to this there will be another set of public announcements from A. The steps after the verification are:

1. Coarse graining of the outcomes.

2. Every party except A_0 and A_1 announces their outcome from the first measurement in $n_{key} \Rightarrow A_0$ and A_1 know the full outcome, everybody else (including E) knows the announced outcomes and the parity of the two unannounced bits.

3. Every party except A_0 and A_2 announces their outcome from the second measurement in $n_{key} \Rightarrow A_0$ and A_2 know the full outcome, everybody else knows the announced outcomes and the parity of the two unannounced bits.

...

N Every party except A_0 and A_{N-1} announces their outcome from the N-th measurement in $n_{key} \Rightarrow A_0$ and A_{N-1} know the full outcome, everybody else knows the announced outcomes and the parity of the two unannounced bits.

At this point each party knows the full set of outcomes for a certain round jointly with A_0, which can act as a "distributor" of information through a set of other announcements. Call $a_{(i,j)}$ the j-th outcome of the i-th party:

$N+1$ A_0 announces $a_{(0,1)} \oplus a_{(2,2)}$. A_1 (which knows $a_{(0,1)}$) can deduce $a_{(2,2)}$, while A_2 (which knows $a_{(2,2)}$ can deduce $a_{(0,1)} \Rightarrow A_0, A_1, A_2$ all know the full set of outcomes for the first and second sets of measurements.

$N+2$ A_0 announces $a_{(0,1)} \oplus a_{(3,3)}$. A_1 and A_2 (which know $a_{(0,1)}$) can deduce $a_{(3,3)}$, while A_3 (which knows $a_{(3,3)}$) can deduce $a_{(0,1)} \Rightarrow A_0, A_1, A_2, A_3$ all know the full set of outcomes for the first set of measurements.

...

repeat iteratively to share the full set of outcomes of the first round with all the parties in the network. At this point the network can choose any one of the bits from the first set of measurements (let us say they choose $a_{(0,1)}$) as the first bit of the raw key.
One thing to be noted is that they can choose only one of the bits from the full set of outcomes from the first round, since the announcements made by A_0 create some correlation which could be exploited by E. To (hopefully) make this clearer I will briefly present the QCKA protocol for 4 parties (A, B, C, D) in Section 4.2.2.1.

The privacy amplification follows the same procedure explained in Section 4.1.4 with the difference that now the LHL (2.2) takes the form (proved in **Lemma 6.4.1** in [34])

$$l \leq \left\lfloor H_{min}^{\varepsilon}(A_0^I|\bar{E}) - \text{leak}_{\text{IR}} - 2\log\left(\frac{1}{2\varepsilon'}\right)\right\rfloor \tag{4.7}$$

where the term leak_{IR} is associated to the leakage introduced by the information reconciliation (IR) part (the announcements made by all the parties). A_0^I represents the (raw) key and $\bar{E}$ is the register of information that E has on top of leak_{IR} (All the details and the explicit proof of security is in Section 4.3.2).

4.2.2.1 $N = 4$ example

let us assume that the verification has already taken place and focus directly on the key generation part of the protocol. One tool that revealed to be useful is what we (informally) called a "knowledge table", shown in Table 4.2. It can be read in this way: in the first column are the announcements of each round (rows) in n_{key}, the bit is announced by the corresponding party (b is announced by B), exception made for the "$\oplus$" announcements, done by A. Each column except the

first represents the knowledge that the party gains at the current round.

Ann.	A	B	C	D	E
$b_0,\ d_0$	(a_0, b_0, c_0, d_0)	$b_0, d_0, a_0 \oplus c_0$	(a_0, b_0, c_0, d_0)	$b_0, d_0, a_0 \oplus c_0$	$b_0, d_0, a_0 \oplus c_0$
$c_1,\ d_1$	(a_1, b_1, c_1, d_1)	(a_1, b_1, c_1, d_1)	$c_1, d_1, a_1 \oplus b_1$	$c_1, d_1, a_1 \oplus b_1$	$c_1, d_1, a_1 \oplus b_1$
$b_2,\ c_2$	(a_2, b_2, c_2, d_2)	$b_2, c_2, a_2 \oplus d_2$	$b_2, c_2, a_2 \oplus d_2$	(a_2, b_2, c_2, d_2)	$b_2, c_2, a_2 \oplus d_2$
$a_0 \oplus b_1$		(a_0, b_0, c_0, d_0)	(a_1, b_1, c_1, d_1)	$a_1 \oplus c_0$	$a_1 \oplus c_0$
$a_0 \oplus d_2$		(a_2, b_2, c_2, d_2)	(a_2, b_2, c_2, d_2)	(a_0, b_0, c_0, d_0)	$a_2 \oplus c_0$

Table 4.2: Taking as an example the first line B and D announce their (coarse grained) outcome, A and C know their own bit $\Rightarrow$ they can deduce the missing bit given the constraint on Δ (there is only one possible string of 4 bits). For the same reason looking at the possible outcomes B, D, and E can only know whether $a_0 = c_0$ or $a_0 = c_0 \oplus 1$.

From this example it is easy to see that by the end of the protocol, all 4 parties know (a_0, b_0, c_0, d_0) and (a_2, b_2, c_2, d_2), so they can all take a_0 as the first bit of K. If they were to take, let us say, both a_0 and a_2, to have 2 bits for the raw key, then E could deduce some information: the chain of announcements $a_2 \oplus c_0 = a_2 \oplus a_1 = a_2 \oplus b_1 = a_2 \oplus a_0$ shows there is a relation between a_0 and a_2, thus using both for K creates a key over which E has more knowledge.

4.3 Security

In this section are the assumptions taken with a brief motivation, and the explicit proof for the QCKA protocol, of which the case $M = 2$ is a particular case.

4.3.1 Assumptions

Following are the assumptions made taken, under which I will prove the security:

1. **Authentication**: The parties share authenticated classical channels.
2. **Network assumption**: The sources are all independent from each other.
3. **IID**: The rounds are independent from each other and are identically distributed.
4. **Asymptotic limit** $n_{test}, n_{key} \to \infty$

Before going through the proof, we can make a few comments about the assumptions. Assumption 1. is made in every QKD proof, even though in this case it is more subtle as local randomness is typically used to create authenticated channels (see Section 4.5). The analogue of Assumption 2. is the free-will or measurement-independence assumption in Bell tests and standard DIQKD, i.e. that the users have sources (of randomness in the case of standard DIQKD) which are uncorrelated from all other sources and hidden variables. Assumption 3. is introduced due to the memory attacks that can arise, in fact as proven in [53], in non-IID conditions it's not possible to self-test mixed states, which is pivotal to the protocol. Assumption 4. allows to work in the asymptotic regime, i.e. with the actual distribution and without needing finite-size corrections. Furthermore this is required because current self-testing proofs in networks without inputs are not robust to noise (see Section 4.5).

4.3.2 Proof

Starting from (4.7) it's possible to address the leakage due to classical post-processing as well. Introducing a change of notation we have

$$l \leq \left\lfloor H_{min}^{\varepsilon}(\bar{X}|\bar{E}) - \text{leak}_{\text{IR}} - 2\log\left(\frac{1}{\varepsilon'}\right)\right\rfloor \tag{4.8}$$

where $\bar{X}$ is the register of the key and $\bar{E}$ is the register of information in the hands of external eavesdropper. The next step is to rewrite $H_{min}^{\varepsilon}(X|E) - \text{leak}_{\text{IR}}$ as $H_{min}^{\varepsilon}(\bar{X}|\bar{\mathscr{C}}\bar{E})$ where $\bar{\mathscr{C}}$ is the information leaked through the announcements for the information reconciliation (IR). Since two parties at a time (let us say A and C) share a secret bit $\bar{\mathscr{C}} = \bar{X} \oplus \bar{C}$.

To bound the smooth-min entropy we use the QAEP in (1.117), or in our updated notation

$$H_{min}^{\varepsilon}(\bar{X}|\bar{\mathscr{C}}\bar{E}) \geq k \cdot H(X|\mathscr{C}E) - \sqrt{k} \cdot \delta(\varepsilon, \eta) \tag{4.9}$$

where k is the length of the raw key, X is one of the bits of the key and C is the other unannounced bit associated to X. Using (4.5) this becomes

$$H_{min}^{\varepsilon}(\bar{X}|\bar{\mathscr{C}}\bar{E}) \geq k \cdot H_{min}(X|\mathscr{C}E) - \sqrt{k} \cdot \delta(\varepsilon, \eta) \tag{4.10}$$

Finally, by means of

$$H_{min}(X|\mathscr{C}E) = -\log p_{guess}$$

we only need to bound the value of the guessing probability. To do so I will extend the proof of Theorem 7 in [9] to 2 parties with 2 sources each, starting by writing the global state of A, C, E

$$\begin{aligned}\varphi_{ACE} &= \rho_{AE_{\alpha_L}E_{\alpha_R}} \otimes \sigma_{CE_{\gamma_L}E_{\gamma_R}} = \\ &= \Big[|0\rangle\langle 0|_A \Big(p_{\alpha_R}(1-p_{\alpha_L})\rho^{01}_{E_{\alpha_L}}\rho^{10}_{E_{\alpha_R}} + p_{\alpha_L}(1-p_{\alpha_R})\rho^{10}_{E_{\alpha_L}}\rho^{01}_{E_{\alpha_R}}\Big) + \\ &+ |1\rangle\langle 1|_A \Big(p_{\alpha_R}p_{\alpha_L}\rho^{10}_{E_{\alpha_L}}\rho^{10}_{E_{\alpha_R}} + (1-p_{\alpha_L})(1-p_{\alpha_R})\rho^{01}_{E_{\alpha_L}}\rho^{01}_{E_{\alpha_R}}\Big)\Big] \otimes \\ &\otimes \Big[|0\rangle\langle 0|_C \Big(p_{\gamma_R}(1-p_{\gamma_L})\sigma^{01}_{E_{\gamma_L}}\sigma^{10}_{E_{\gamma_R}} + p_{\gamma_L}(1-p_{\gamma_R})\sigma^{10}_{E_{\gamma_L}}\sigma^{01}_{E_{\gamma_R}}\Big) + \\ &+ |1\rangle\langle 1|_C \Big(p_{\gamma_R}p_{\gamma_L}\sigma^{10}_{E_{\gamma_L}}\sigma^{10}_{E_{\gamma_R}} + (1-p_{\gamma_L})(1-p_{\gamma_R})\sigma^{01}_{E_{\gamma_L}}\sigma^{01}_{E_{\gamma_R}}\Big)\Big]\end{aligned} \tag{4.11}$$

p_i is the probability that the source i sends the token to their left (so given a party with two sources, i_L to the left and i_R to the right, i_L sends the token away from the party and i_R towards the party). ρ^{01} and ρ^{10} (and similarly for σ) represent the state in possess of E when the token arrive to A (C for σ), respectively from the right and left sources.
The states for A and C are the coarse grained 0 and 1, since the 0 is associated to probability of receiving one token from both sources (fine-grained 2) plus probability of receiving no token from both sources (fine-grained 0). In the same way the state $|1\rangle\langle 1|$ is associated to receiving one token from one of the sources (fine-grained 1_L and 1_R).
The factorizations are possible because we assume the parties and the sources to be independent from each other.

We can explicitly find the global state calculating all the products:

$$
\begin{aligned}
&|0\rangle\langle 0|_A \otimes \big[\, |0\rangle\langle 0|_{A\oplus C} p_{\alpha_R}(1-p_{\alpha_L})p_{\gamma_R}(1-p_{\gamma_L})\rho^{0110}\sigma^{0110} + \\
&+ |0\rangle\langle 0|_{A\oplus C} p_{\alpha_R}(1-p_{\alpha_L})p_{\gamma_L}(1-p_{\gamma_R})\rho^{0110}\sigma^{1001} + \\
&+ |0\rangle\langle 0|_{A\oplus C} p_{\alpha_L}(1-p_{\alpha_R})p_{\gamma_R}(1-p_{\gamma_L})\rho^{1001}\sigma^{0110} + \\
&+ |0\rangle\langle 0|_{A\oplus C} p_{\alpha_L}(1-p_{\alpha_R})p_{\gamma_L}(1-p_{\gamma_R})\rho^{1001}\sigma^{1001} + \\
&+ |1\rangle\langle 1|_{A\oplus C} p_{\alpha_R}(1-p_{\alpha_L})p_{\gamma_R}p_{\gamma_L}\rho^{0110}\sigma^{1010} + \\
&+ |1\rangle\langle 1|_{A\oplus C} p_{\alpha_R}(1-p_{\alpha_L})(1-p_{\gamma_L})(1-p_{\gamma_R})\rho^{0110}\sigma^{0101} + \\
&+ |1\rangle\langle 1|_{A\oplus C} p_{\alpha_L}(1-p_{\alpha_R})p_{\gamma_R}p_{\gamma_L}\rho^{1001}\sigma^{1010} + \\
&+ |1\rangle\langle 1|_{A\oplus C} p_{\alpha_L}(1-p_{\alpha_R})(1-p_{\gamma_L})(1-p_{\gamma_R})\rho^{1001}\sigma^{0101}\big] + \\
&+ |1\rangle\langle 1|_A \otimes \big[\, |0\rangle\langle 0|_{A\oplus C} p_{\alpha_R}p_{\alpha_L}p_{\gamma_R}p_{\gamma_L}\rho^{1010}\sigma^{1010} + \\
&+ |0\rangle\langle 0|_{A\oplus C} p_{\alpha_R}p_{\alpha_L}(1-p_{\gamma_L})(1-p_{\gamma_R})\rho^{1010}\sigma^{0101} + \\
&+ |0\rangle\langle 0|_{A\oplus C} (1-p_{\alpha_R})(1-p_{\alpha_L})p_{\gamma_R}p_{\gamma_L}\rho^{0101}\sigma^{1010} + \\
&+ |0\rangle\langle 0|_{A\oplus C} (1-p_{\alpha_R})(1-p_{\alpha_L})(1-p_{\gamma_L})(1-p_{\gamma_R})\rho^{0101}\sigma^{0101} + \\
&+ |1\rangle\langle 1|_{A\oplus C} p_{\alpha_R}p_{\alpha_L}p_{\gamma_R}(1-p_{\gamma_L})\rho^{1010}\sigma^{0110} + \\
&+ |1\rangle\langle 1|_{A\oplus C} p_{\alpha_R}p_{\alpha_L}p_{\gamma_L}(1-p_{\gamma_R})\rho^{1010}\sigma^{1001} + \\
&+ |1\rangle\langle 1|_{A\oplus C} (1-p_{\alpha_R})(1-p_{\alpha_L})p_{\gamma_R}(1-p_{\gamma_L})\rho^{0101}\sigma^{0110} + \\
&+ |1\rangle\langle 1|_{A\oplus C} (1-p_{\alpha_R})(1-p_{\alpha_L})p_{\gamma_L}(1-p_{\gamma_R})\rho^{0101}\sigma^{1001}\big]
\end{aligned}
\tag{4.12}
$$

where $\rho^{0110} = \rho^{01}_{E_{\alpha_R}}\rho^{10}_{E_{\alpha_L}}$ and similarly for all the others. Furthermore I grouped the product state of A and C with the coarse graining $A\oplus C$, so $|0\rangle\langle 0|_{A\oplus C} = |0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_C + |1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_C$ and $|1\rangle\langle 1|_{A\oplus C} = |0\rangle\langle 0|_A \otimes |1\rangle\langle 1|_C + |1\rangle\langle 1|_A \otimes |0\rangle\langle 0|_C$, in order to make the next step easier to see.

Focusing on A and their sources only (α_L and α_R) we can notice that the probabilities in the square brackets of the first term ($A = 0$) always appear as $p_{\alpha_i}(1-p_{\alpha_j})$ with $i = L, R$ and $j \neq i$. This reflects the fact that A gets either 0 or 2 tokens from the sources, which will be sent respectively away or towards A. Calling $\bar{z} = b_1b_2b_3b_4$ the string on top of ρ, the coefficient in front of $\rho^{\bar{z}}$ is $p^{b_1}_{\alpha_L}(1-p_{\alpha_L})^{b_2}p^{b_3}_{\alpha_R}(1-p_{\alpha_R})^{b_4}$. These numbers are subject to the constraints $b_2 = 1-b_1$ and $b_4 = 1-b_3$, since each source only distributes 1 token, thus we can rename them for clarity $b_1 = s_1$, $b_2 = 1-s_1$, $b_3 = s_2$ and $b_4 = 1-s_2$. The coefficients in the square bracket associated to $A = 0$ can then be rewritten as a product of Bernoulli distributions, specifically

$$
Q = \sum_{s_1,s_2=1-s_1,s_3,s_4\in\{0,1\}} p^{s_1}_{\alpha_L}(1-p_{\alpha_L})^{1-s_1}p^{s_2}_{\alpha_R}(1-p_{\alpha_R})^{1-s_2}p^{s_3}_{\gamma_L}(1-p_{\gamma_L})^{1-s_3}p^{s_4}_{\gamma_R}(1-p_{\gamma_R})^{1-s_4}
\tag{4.13}
$$

with s_1, s_2 associated to A, while s_3 and s_4 are the same thing for C. We also have $s_2 = 1 - s_1$ to enforce the fact that the sources adjacent to A do opposite things. Before calculating the corresponding term for $A = 1$ we can notice that given a string of 4 bits there are $2^4 = 16$ possible combinations, and the constraint $s_2 = 1 - s_1$ removes half of them. Consequently $s_2 = s_1$ takes the other half, meaning that the sum of the two must be 1 and we can write them as Q and $1 - Q$.

We can finally rewrite the global state as

$$\phi_{A,A\oplus C,E} = Q \left|0\right\rangle \left\langle 0\right|_A \otimes \tau^0_{A\oplus C,E} + (1-Q) \left|1\right\rangle \left\langle 1\right|_A \otimes \tau^1_{A\oplus C,E} \tag{4.14}$$

where

$$\begin{aligned}
\tau^0_{A\oplus C,E} &= \frac{\left|0\right\rangle \left\langle 0\right|_{A\oplus C} \otimes \rho^0_E \sigma^0_E + \left|1\right\rangle \left\langle 1\right|_{A\oplus C} \otimes \rho^0_E \sigma^1_E}{Q} \\
\tau^1_{A\oplus C,E} &= \frac{\left|0\right\rangle \left\langle 0\right|_{A\oplus C} \otimes \rho^1_E \sigma^1_E + \left|1\right\rangle \left\langle 1\right|_{A\oplus C} \otimes \rho^1_E \sigma^0_E}{1-Q} \\
\rho^0_E &= p_{\alpha_R}(1-p_{\alpha_L})\rho^{01}_{E_{\alpha_L}}\rho^{10}_{E_{\alpha_R}} + p_{\alpha_L}(1-p_{\alpha_R})\rho^{10}_{E_{\alpha_L}}\rho^{01}_{E_{\alpha_R}} \\
\rho^1_E &= p_{\alpha_R}p_{\alpha_L}\rho^{10}_{E_{\alpha_L}}\rho^{10}_{E_{\alpha_R}} + (1-p_{\alpha_L})(1-p_{\alpha_R})\rho^{01}_{E_{\alpha_L}}\rho^{01}_{E_{\alpha_R}} \\
\sigma^0_E &= p_{\gamma_R}(1-p_{\gamma_L})\rho^{01}_{E_{\gamma_L}}\rho^{10}_{E_{\gamma_R}} + p_{\gamma_L}(1-p_{\gamma_R})\rho^{10}_{E_{\gamma_L}}\rho^{01}_{E_{\gamma_R}} \\
\sigma^1_E &= p_{\gamma_R}p_{\gamma_L}\rho^{10}_{E_{\gamma_L}}\rho^{10}_{E_{\gamma_R}} + (1-p_{\gamma_L})(1-p_{\gamma_R})\rho^{01}_{E_{\gamma_L}}\rho^{01}_{E_{\gamma_R}}
\end{aligned} \tag{4.15}$$

The point of this long and verbose manipulation is that the guessing probability for the state $\phi_{A,A\oplus C,E}$ is known, and by Holevo-Helstrom (Theorem 3.4 [51]) we have

$$p_g(A|A\oplus C,E) \leq \frac{1}{2}\left(1 + \left|\left|Q\tau^0_{A\oplus C,E} - (1-Q)\tau^1_{A\oplus C,E}\right|\right|_1\right) \tag{4.16}$$

Focusing on the object inside the trace distance, it can be manipulated further

$$\begin{aligned}
&\left|\left|\left|0\right\rangle \left\langle 0\right|_{A\oplus C} (\rho^0_E\sigma^0_E - \rho^1_E\sigma^1_E) + \left|1\right\rangle \left\langle 1\right|_{A\oplus C} (\rho^0_E\sigma^1_E - \rho^1_E\sigma^0_E)\right|\right|_1 \leq \\
&\leq 1 \cdot \left|\left|\rho^0_E\sigma^0_E - \rho^1_E\sigma^1_E\right|\right|_1 + 1 \cdot \left|\left|\rho^0_E\sigma^1_E - \rho^1_E\sigma^0_E\right|\right|_1
\end{aligned} \tag{4.17}$$

where I used the triangle inequality ([30] Eq(9.32) to Eq(9.34)) and the property of the trace $||\rho\otimes\sigma||_1 = ||\rho||_1 \cdot ||\sigma||_1$. The first term on the right-hand side of the previous equation can be explicited and is

$$\Big|\Big|(1-p_{\alpha_L})(1-p_{\gamma_L})\rho^{01}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_L}}\Big[p_{\alpha_R}p_{\gamma_R}\rho^{10}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_R}}-(1-p_{\alpha_R})(1-p_{\gamma_L})\rho^{01}_{E_{\alpha_R}}\rho^{01}_{E_{\gamma_R}}\Big]+$$
$$+p_{\alpha_R}p_{\gamma_L}\rho^{10}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_L}}\Big[(1-p_{\alpha_L})(1-p_{\gamma_R})\rho^{01}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_R}}-p_{\alpha_L}p_{\gamma_R}\rho^{10}_{E_{\alpha_L}}\rho^{10}_{E_{\gamma_R}}\Big]+ \tag{4.18}$$
$$+(1-p_{\alpha_R})p_{\gamma_R}\rho^{01}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_R}}\Big[p_{\alpha_L}(1-p_{\gamma_L})\rho^{10}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_L}}-(1-p_{\alpha_L})p_{\gamma_L}\rho^{01}_{E_{\alpha_L}}\rho^{10}_{E_{\gamma_L}}\Big]+$$
$$+p_{\alpha_L}(1-p_{\gamma_R})\rho^{10}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_R}}\Big[(1-p_{\alpha_R})p_{\gamma_L}\rho^{01}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_L}}-p_{\gamma_R}(1-p_{\alpha_R})\rho^{10}_{E_{\alpha_R}}\rho^{01}_{E_{\gamma_L}}\Big]\Big|\Big|_1$$

thus using again the triangle inequality and the linearity of the trace distance we have

$$\big|\big|\rho^0_E\sigma^0_E-\rho^1_e\sigma^1_E\big|\big|_1\leq \tag{4.19}$$
$$\leq\Big|(1-p_{\alpha_L})(1-p_{\gamma_L})\Big|\cdot\Big|\Big|p_{\alpha_R}p_{\gamma_R}\rho^{10}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_R}}-(1-p_{\alpha_R})(1-p_{\gamma_L})\rho^{01}_{E_{\alpha_R}}\rho^{01}_{E_{\gamma_R}}\Big|\Big|_1+$$
$$+\Big|p_{\alpha_R}p_{\gamma_L}\rho^{10}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_L}}\Big|\cdot\Big|\Big|(1-p_{\alpha_L})(1-p_{\gamma_R})\rho^{01}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_R}}-p_{\alpha_L}p_{\gamma_R}\rho^{10}_{E_{\alpha_L}}\rho^{10}_{E_{\gamma_R}}\Big|\Big|_1+$$
$$+\Big|(1-p_{\alpha_R})p_{\gamma_R}\rho^{01}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_R}}\Big|\cdot\Big|\Big|p_{\alpha_L}(1-p_{\gamma_L})\rho^{10}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_L}}-(1-p_{\alpha_L})p_{\gamma_L}\rho^{01}_{E_{\alpha_L}}\rho^{10}_{E_{\gamma_L}}\Big|\Big|_1+$$
$$+\Big|p_{\alpha_L}(1-p_{\gamma_R})\rho^{10}_{E_{\alpha_L}}\rho^{01}_{E_{\gamma_R}}\Big|\cdot\Big|\Big|(1-p_{\alpha_R})p_{\gamma_L}\rho^{01}_{E_{\alpha_R}}\rho^{10}_{E_{\gamma_L}}-p_{\gamma_R}(1-p_{\alpha_R})\rho^{10}_{E_{\alpha_R}}\rho^{01}_{E_{\gamma_L}}\Big|\Big|_1$$

The same can be done to the term $||\rho^0_E\sigma^1_E-\rho^1_E\sigma^0_E||_1$, obtaining a similar result. At this point, we can notice that for any ρ and σ

$$||p\rho-q\sigma||_1=\Big|\Big|2\frac{p}{2}\rho-2\frac{q}{2}\sigma\Big|\Big|_1=\Big|\Big|\frac{p-q}{2}(\rho+\sigma)+\frac{p+q}{2}(\rho-\sigma)\Big|\Big|_1\leq$$
$$\leq\frac{p+q}{2}||\rho-\sigma||_1+\frac{|p-q|}{2}||\rho+\sigma||_1\leq(p+q)\sqrt{1-F(\rho,\sigma)}+|p-q| \tag{4.20}$$

where $F(\rho,\sigma)$ is the fidelity. Calling $F_{\alpha_{R(L)}}=F\left(\rho^{10}_{E_{\alpha_{R(L)}}},\rho^{01}_{E_{\alpha_{R(L)}}}\right)$ and $F_{\gamma_{R(L)}}=F\left(\sigma^{10}_{E_{\gamma_{R(L)}}},\sigma^{01}_{E_{\gamma_{R(L)}}}\right)$ and using the above equation on all terms of (4.19) we end up with

$$
\begin{aligned}
&p_g(A|A\oplus C,E)\leq \qquad (4.21)\\
&\leq \frac{1}{2}\Big(1+\big|(1-p_{\alpha_L})(1-p_{\gamma_L})\big|\left[(1-p_{\alpha_R}-p_{\gamma_R}+2p_{\alpha_R}p_{\gamma_R})\sqrt{1-F_{\alpha_R}F_{\gamma_R}}+\big|1-p_{\alpha_R}-p_{\gamma_R}\big|\right]+\\
&+\big|p_{\alpha_R}p_{\gamma_L}\big|\left[(1-p_{\alpha_L}-p_{\gamma_R}+2p_{\alpha_L}p_{\gamma_R})\sqrt{1-F_{\alpha_L}F_{\gamma_R}}+\big|1-p_{\alpha_L}-p_{\gamma_R}\big|\right]+\\
&+\big|(1-p_{\alpha_R})p_{\gamma_R}\big|\left[(p_{\alpha_L}+p_{\gamma_L}-2p_{\alpha_L}p_{\gamma_L})\sqrt{1-F_{\alpha_L}F_{\gamma_L}}+\big|p_{\alpha_L}-p_{\gamma_L}\big|\right]+\\
&+\big|p_{\alpha_L}(1-p_{\gamma_R})\big|\left[(p_{\alpha_R}+p_{\gamma_L}-2p_{\alpha_R}p_{\gamma_L})\sqrt{1-F_{\alpha_R}F_{\gamma_L}}+\big|p_{\alpha_R}-p_{\gamma_L}\big|\right]+\\
&+\big|(1-p_{\alpha_L})p_{\gamma_L}\big|\left[(1-p_{\alpha_R}-p_{\gamma_R}+2p_{\alpha_R}p_{\gamma_R})\sqrt{1-F_{\alpha_R}F_{\gamma_R}}+\big|1-p_{\alpha_R}-p_{\gamma_R}\big|\right]+\\
&+\big|p_{\alpha_R}(1-p_{\gamma_L})\big|\left[(1-p_{\alpha_L}-p_{\gamma_R}+2p_{\alpha_L}p_{\gamma_R})\sqrt{1-F_{\alpha_L}F_{\gamma_R}}+\big|1-p_{\alpha_L}-p_{\gamma_R}\big|\right]+\\
&+\big|(1-p_{\alpha_R})p_{\gamma_R}\big|\left[(1-p_{\alpha_L}-p_{\gamma_L}+2p_{\alpha_L}p_{\gamma_L})\sqrt{1-F_{\alpha_L}F_{\gamma_L}}+\big|1-p_{\alpha_L}-p_{\gamma_L}\big|\right]+\\
&+\big|p_{\alpha_L}(1-p_{\gamma_R})\big|\left[(1-p_{\alpha_R}-p_{\gamma_L}+2p_{\alpha_R}p_{\gamma_L})\sqrt{1-F_{\alpha_R}F_{\gamma_L}}+\big|1-p_{\alpha_R}-p_{\gamma_L}\big|\right]\Big)
\end{aligned}
$$

This can be simplified into

$$
\begin{aligned}
&p_g(A|A\oplus C,E)\leq \qquad (4.22)\\
&\leq \frac{1}{2}\Big(1+(1-p_{\alpha_L})\left[(1-p_{\alpha_R}-p_{\gamma_R}+2p_{\alpha_R}p_{\gamma_R})\sqrt{1-F_{\alpha_R}F_{\gamma_R}}+\big|1-p_{\alpha_R}-p_{\gamma_R}\big|\right]+\\
&+p_{\alpha_R}\left[(1-p_{\alpha_L}-p_{\gamma_R}+2p_{\alpha_L}p_{\gamma_R})\sqrt{1-F_{\alpha_L}F_{\gamma_R}}+\big|1-p_{\alpha_L}-p_{\gamma_R}\big|\right]+\\
&+\big|(1-p_{\alpha_R})p_{\gamma_R}\big|\left[\sqrt{1-F_{\alpha_L}F_{\gamma_L}}+\big|p_{\alpha_L}-p_{\gamma_L}\big|+\big|1-p_{\alpha_L}-p_{\gamma_L}\big|\right]+\\
&+\big|p_{\alpha_L}(1-p_{\gamma_R})\big|\left[\sqrt{1-F_{\alpha_R}F_{\gamma_L}}+\big|p_{\alpha_R}-p_{\gamma_L}\big|+\big|1-p_{\alpha_R}-p_{\gamma_L}\big|\right]\Big)
\end{aligned}
$$

which finally yields, in the case $p_{\alpha_L}=p_{\alpha_R}=p_{\gamma_L}=p_{\gamma_R}=\frac{1}{2}$, our result

$$
\begin{aligned}
&p_g(A|A\oplus C,E)\leq \qquad (4.23)\\
&\leq \frac{1}{2}\left[1+\frac{1}{4}\left(\sqrt{1-F_{\alpha_R}F_{\gamma_R}}+\sqrt{1-F_{\alpha_L}F_{\gamma_R}}+\sqrt{1-F_{\alpha_L}F_{\gamma_L}}+\sqrt{1-F_{\alpha_R}F_{\gamma_L}}\right)\right]
\end{aligned}
$$

This equation respects the constraint

$$
F_{\alpha_R}F_{\alpha_L}F_{\gamma_R}F_{\gamma_L}\geq R \qquad (4.24)
$$

with R known as the *global coherence factor*, introduced in Appendix B.2 (specifically in Theorem 6) of [9]. Given its definition in the form of Eq.(B51) of [9]

$$R \leq \left|\left|\mathrm{tr}_{E_\xi} \left|01, j_\xi^{01}\right\rangle \left\langle 10, j_\xi^{10}\right|\right|\right|_1 \quad \forall \xi (\equiv \text{sources}) \tag{4.25}$$

it's easy to see that (following Eq.(B52), [9]) we get the inequality (4.24).
Now we can analitically find the maximum for $p_g(A|A \oplus C, E)$: consider the function f given by (4.23), ignoring the translation and rescaling factors, suppose the maximum is attained at

$$\max f = f(F^*_{\alpha_R}, F^*_{\alpha_L}, F^*_{\gamma_R}, F^*_{\gamma_L}) \tag{4.26}$$

and assume $F^*_{\alpha_R} < F^*_{\alpha_L}$. Now consider the value of the function at a slightly perturbed point:

$$\begin{aligned} &f(F^*_{\alpha_R} + \delta, F^*_{\alpha_L} - \delta, F^*_{\gamma_R}, F^*_{\gamma_L}) = \\ &= \sqrt{1 - (F^*_{\alpha_R} + \delta)F^*_{\gamma_R}} + \sqrt{1 - (F^*_{\alpha_L} - \delta)F^*_{\gamma_R}} + \sqrt{1 - F^*_{\alpha_L}F^*_{\gamma_L}} + \sqrt{1 - F^*_{\alpha_R}F^*_{\gamma_L}} \\ &= \sqrt{1 - F^*_{\alpha_R}F^*_{\gamma_R} - \delta F^*_{\gamma_R}} + \sqrt{1 - F^*_{\alpha_L}F^*_{\gamma_R} + \delta F^*_{\gamma_R}} + \sqrt{1 - F^*_{\alpha_L}F^*_{\gamma_L}} + \sqrt{1 - F^*_{\alpha_R}F^*_{\gamma_L}} \end{aligned} \tag{4.27}$$

It holds that $\sqrt{a - \delta} + \sqrt{b + \delta} > \sqrt{a} + \sqrt{b}$ for $a > b$ and δ small enough, which can be verified by examining the Taylor expansion of the square root.

$$\sqrt{a - x} + \sqrt{b + x} \overset{T}{\sim} \sqrt{a} - \frac{x}{2a} + \sqrt{b} + \frac{x}{2b}\bigg|_{a>b} > \sqrt{a} + \sqrt{b} \tag{4.28}$$

Since $F^*_{\alpha_R} < F^*_{\alpha_L}$ implies

$$1 - F^*_{\alpha_R}F^*_{\gamma_{R/L}} > 1 - F^*_{\alpha_L}F^*_{\gamma_{R/L}}, \tag{4.29}$$

it follows that

$$f(F^*_{\alpha_R} + \delta, F^*_{\alpha_L} - \delta, F^*_{\gamma_R}, F^*_{\gamma_L}) > f(F^*_{\alpha_R}, F^*_{\alpha_L}, F^*_{\gamma_R}, F^*_{\gamma_L}), \tag{4.30}$$

which contradicts the assumption that the latter point was optimal.
We still need to verify that this new point lies within the domain. It is easy to see that for δ small enough, the fidelities still belong to $[0, 1]$, and moreover, the constraint $F_{\alpha_R}F_{\alpha_L}F_{\gamma_R}F_{\gamma_L} \geq R$ is also satisfied:

$$\begin{aligned} (F^*_{\alpha_R} + \delta)(F^*_{\alpha_L} - \delta)F^*_{\gamma_R}F^*_{\gamma_L} = {} & F^*_{\alpha_R}F^*_{\alpha_L}F^*_{\gamma_R}F^*_{\gamma_L} + \delta F^*_{\alpha_L}F^*_{\gamma_R}F^*_{\gamma_L} \\ & - \delta F^*_{\alpha_R}F^*_{\gamma_R}F^*_{\gamma_L} - \delta^2 F^*_{\gamma_R}F^*_{\gamma_L} \geq R. \end{aligned} \tag{4.31}$$

Given the symmetry of the function with respect to swapping F_{α_R} and F_{α_L}, the same contradiction arises if $F^*_{\alpha_R} > F^*_{\alpha_L}$. A similar argument applies to $F^*_{\gamma_R}$ and $F^*_{\gamma_L}$, therefore the maximum must occur at

$$F^*_{\alpha_R} = F^*_{\alpha_L} \quad \text{and} \quad F^*_{\gamma_R} = F^*_{\gamma_L}. \tag{4.32}$$

Since the function is continuous and defined over a compact domain, a maximum must exist, and from the above symmetry argument, it must occur at

$$f(F^*_{\alpha_R} = F^*_{\alpha_L}, F^*_{\gamma_R} = F^*_{\gamma_L}) = 4\sqrt{1 - F^*_{\alpha_R} F^*_{\gamma_R}}, \tag{4.33}$$

under the constraint that $(F^*_{\alpha_R} F^*_{\gamma_R})^2 \geq R$. From this, it follows that the maximum value is $\max f = 4\sqrt{1-\sqrt{R}}$, meaning that the complete expression for the min-entropy is

$$H_{min}(A|\mathscr{C}E) \geq -\log\left[\frac{1}{2}\left(1+\sqrt{1-\sqrt{R}}\right)\right] = 1 - \log\left(1+\sqrt{1-\sqrt{R}}\right) \tag{4.34}$$

□

To get an idea of how this protocol performs we can use numerical results from Fig. 9 of [9], which indicates that for values of $u \approx 0.8813$ we have a von Neumann entropy $H(A|E)$ of ≈ 0.274. This can be used in the following equation (obtained in [9]) to find a value for R

$$H(A|E) \geq 1 - h\left(\frac{1+R}{2}\right) + H\left(\frac{\left(1+\sqrt{R}\right)^2}{4}, \frac{\left(1-\sqrt{R}\right)^2}{4}, \frac{1-R}{4}, \frac{1-R}{4}\right) \tag{4.35}$$

where

$$h(p) = -p\log p - (1-p)\log(1-p), \ , \quad H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \tag{4.36}$$

with $\sum_i p_i = 1, p_i \geq 0 \ \forall i$, to get $R \approx 0.2032$. Finally using (4.34) we find $H_{min}(A|\mathscr{C}E) \approx 0.2000$, and selecting the distance from uniform distribution ε' and the smoothing radius ε, one can get bounds on both the smooth min-entropy (1.117) and the leftover hash lemma (2.2).

4.4 Untrusted network

In this section I will show how the protocol can be extended to the case of a (mostly) untrusted network.
Assume now that A and B still want to share a secret key but they do not trust the rest of the network except for their respective right and left adjacent parties. The network is shown in Fig. 4.3. The setup is such because the objective is to take advantage of the result in [9], where the authors prove that the lower bound in the randomness of the outcomes and the other results of [41], are topologically robust. This means that it is not necessary to know the topology of the network except for those few sources and parties shown to be outside of the cloud. Based on this, each party that wants to share a key must trust only the adjacent parties ("bodyguards") of everyone who wants to participate, plus the two connecting sources.

Figure 4.3: In the gray cloud is the rest of the parties, which could collaborate leaking information to steal the key that A (left with the keylock) and B (right with the keylock) are creating. Both A and B trust all the parties (circles) and the sources (stars) outside the cloud (pink).

In this scenario, things are more complicated: because of the collaboration of malicious parties, we can in fact assume that they will share everything that they can, in particular the fine grained ($\{0, 1_L, 1_R, 2\}$) outcomes of each. Looking back at Section 4.1.3 and how the key is generated, if the network was to proceed as

usual everybody would need to announce their coarse grained ($\{0, 1\}$) outcomes: this would result in a certain and perfect leakage of the key!
To see this consider A, their bodyguards, and the two malicious parties adjacent to the bodyguards (E_L on the left of the leftmost bodyguard A_L, and E_R on the right of the rightmost bodyguard A_R). Before continuing it is important to notice that, based on how the network was built, we have some constraints on the fine-grained outcomes: specifically there cannott be multiple (fine-grained) 2s or 0s "quasi-adjacent", since this would mean that some sources send out the wrong number of tokens, and the total number of tokens is fixed by the number of sources. We use the "quasi-adjacent" nomenclature to indicate that there can be any number of 1s between two numbers which are not 1s: for example the string of outcomes 2112 is not allowed because the first number different from 1 must be a 0.
At this point assume that the fine-grained outcome of E_L and E_R is the same and is equal to 2, the possible scenarios are:

1. $A_L = A_R = 0^c$ (their coarse grained announcement), then E_L and E_R know that those are 0^f (fine-grained), since adjacent 2s or 0s are not allowed, and for the same reason they can deduce that A has 2^f (see Fig. 4.4).

2. $A_L = A_R = 1^c$, then E_L and E_R know that those are 1^f (does not matter if 1_L or 1_R), thus they can deduce that A has 0^f.

3. $A_L = 1^c, A_R = 0^c$, then E_L and E_R know that $A_L = 1^f$ and $A_R = 0^f$, thus $A = 1^f$. The same goes for $A_R = 1^c, A_L = 0^c$.

The case where $E_L = E_R = 0^f$ is identical to this, inverting 2s and 0s one gets the same results because of the symmetry associated to the coarse graining. We can quickly show that the problem persists if $E_L = 2^f$ and $E_R = 0^f$:

4. $A_L = A_R = 0^c$, then E_L and E_R know that $A_L = 0^f$ and $A_R = 2^f$, thus A has 1^f.

5. $A_L = A_R = 1^c$, then A has 1^f.

6. $A_L = 1^c, A_R = 0^c$, then E_L and E_R know that $A_L = 1^f$ and $A_R = 2^f$, thus $A = 0^f$, while for $A_R = 1^c, A_L = 0^c$ they can deduce $A_L = 0^f, A_R = 1^f$ and finally $A = 2^f$.

and a similar result is obtained for $E_L = 0^f, E_R = 2^f$ The remaining cases, where either one of the Es outputs a 1_i^f, can be ignored and reduced to one of the previous cases since the relevant numbers are those different from 1.

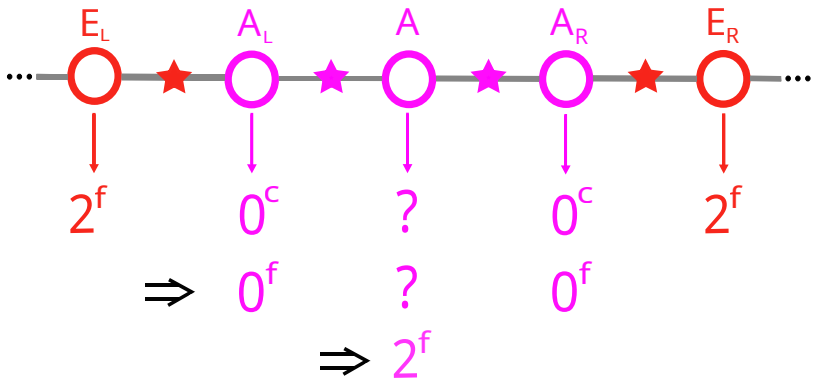

Figure 4.4: In red are the malicious parties (circles) and sources (stars), while in pink the trusted, each with their own outcomes. The scenario is to be interpreted from E_L's and E_R's point of view, where they share with each other their fine grained outcomes (2^f), and have access to the announced coarse grained outcomes (0^c for both A_L and A_R), while A's outcome is (for now) unknown. Since each source emits exactly one token there cannot be adjacent 2s or 0s, with this information E_L and E_R can deduce that the 0^c of A_L and A_R are actually 0^fs (second line). With the same reasoning they can deduce that A's outcome is 2^f (last line), and thus that the corresponding bit of the secret key is a 0 (due to the coarse graining).

The only case (that we know of) where it's possible to implement the protocol in a network such as in Fig. 4.3, is if every party between A and B outputs 1^c, in fact if this was the case there would be no 0^f and 2^f to use to infer the outcome of A. However aborting the protocol until the whole malicious network outputs 1s only would vastly increase the cost of creating a secret key.

Two other solutions are studied in the next section.

4.5 Considerations and outlook

In this last paragraph are some final considerations and open questions.

- **Solutions for the untrusted network:** Here we present two obvious ways to address the problem presented at the end of Section 4.4. The first one would be to remove the untrusted parties in between A and B, the second one to make one of the two bodyguards silent for the whole duration of the protocol. Let us analyze them one by one, starting from this las one:

  1. **Silent bodyguard.** Assume $E_L = E_R = 2^f$ and that the outcome of A_L was unannounced. Then for the malicious network it would be equally probable to have $A_L = 0^c$ and $A = 2^f$ or $A_L = 1^c$ and $A = 1^f$, i.e. both case 1 and 3 in Section 4.4 would be possible. However this solution makes (allegedly) impossible the information reconciliation subroutine of Section 4.2.2, since B does not know the outcomes of 3 parties (A, A_L, B_L) instead of 1 (A). To address this, we could imagine that A and B have control over A_L and B_L respectively, behaving as a single party, for example in a single laboratory. The proof in Section 4.3 would probably need to be partially reworked.
  2. **Entanglement swapping.** At each iteration of the protocol the parties could require that the malicious networks performs an entanglement swapping protocol, as to connect A_L with B_R and A_R with B_L. In this way the malicious network is effectively cut out of the protocol, allowing the trusted parties to create a subnetwork where they could perform the protocol as presented (see Fig. 4.5). To check that the malicious parties actually do the entanglement swapping, the parties could perform some CHSH test to check that they actually share two entangled particles, at the price of a higher number of measurements.

- **Scaling:** One important point of the protocol is that, in the QKD case (Section 4.1 and Section 4.4), it scales linearly with the number of parties interested in sharing the key, rather than the number of parties in the whole network. To see this we can look at Table 4.2 and check that to share the a bit between two parties, one measurement is enough as A and C both know (a_0, b_0, c_0, d_0). Thus m measurement yield a raw key of length m.
  Similarly to share a bit between four parties, each one of them has to do three measurements, meaning that to have a raw key of length m each has to do $3m$ measurements. Calling $N = n_{rand} + n_{test} + n_{key}$ and M the number of parties who want to share the key, the number of measurements necessary is $O(N \cdot M)$.

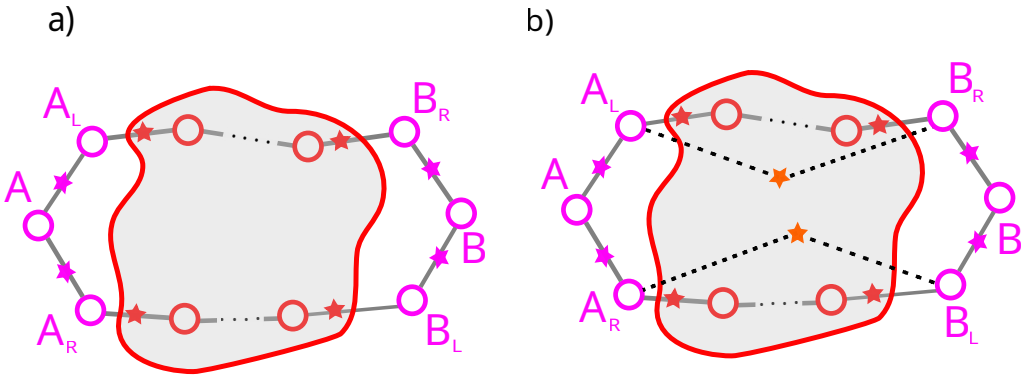

Figure 4.5: In red are the malicious parties (circles) and sources (stars), while in pink are the trusted parties and sources. **a)** The untrusted network can take any configuration, so for simplicity we keep the ring topology. **b)** After the entanglement swapping performed by the malicious parties, A_L effectively shares a new source with B_R (orange star and dotted line), and the same goes for A_R with B_L. In this way the trusted parties can exclude the malicious parties from the protocol.

The length of the raw key will be $|K_{raw}| = n_{key}$, while in the best case scenario ((1.117) and (4.7) are saturated), the length of the final key as a function of all the other variables is

$$\begin{aligned} \left|K(n_{key}, R, \varepsilon_{PA}, \varepsilon_{\text{smooth}}, \eta)\right| &= n_{key}\left[1 - \log\left(1 + \sqrt{1 - \sqrt{R}}\right)\right] - \\ &- \sqrt{n_{key}}\delta(\varepsilon_{\text{smooth}}, \eta) - 2\log\left(\frac{1}{2\varepsilon_{PA}}\right) \end{aligned} \tag{4.37}$$

where ε_{PA} is the distance from uniform distribution, $\varepsilon_{\text{smooth}}$ is the smoothing parameter of the entropy, R is the global coherence defined in [9], η is explicitly in (1.120), and $\delta(\varepsilon_{\text{smooth}}, \eta) = 4\log\eta\sqrt{\log(2/\varepsilon^2_{\text{smooth}})}$ for $n_{key} \geq \frac{8}{5}\log(2/\varepsilon^2_{\text{smooth}})$.

- **Assumptions:**

  1. *Authentication.* Having authenticated channels is necessary to be sure of the party we are communicating with, and to avoid man-in-the-middle or intercept-resend attacks (see Section 2.1.2). This is a well-studied problem in classical cryptography, and most schemes rely on Alice having some seed randomness to generate public-private key pairs. One possible way to address this could be to use a short OTP (one-time-pad) to establish a first communication, and then use some of the randomness generated in the "Verification" part of the protocol (Section 4.1.2) to create the authenticated channel. This is a standard assumption in any QKD protocol and does not take away from the results of Chapter 4, which focus on removing the local randomness of the parties.
  2. *Noise-robustness.* As mentioned in Section 4.3.1, the presence of noise is disruptive to the protocol. In [10] the authors prove that the nonlocality of a special family of distributions, of which the RGB4 is a member, is robust to arbitrary types of noise (with different tolerances). However this is not true for self-testing results, which are not proven in the presence of noise, thus requiring to assume no noise in the setup.
  3. *Infinite rounds.* This assumption is partially related to the point above, in fact in a finite number of rounds it is impossible to have the resulting distribution to match perfectly the RGB4. This can be interpreted as a kind of noise in sources distributing the states.
     Furthermore this allows to neglect finite-size corrections.
  4. *IID.* The *identical and independent distribution* assumption implies that there is no correlation between the states distributed by the sources, and that each round the state distributed is $|\psi^+\rangle$. We take this assumption as a consequence of the results of [53], where the authors prove two main results: binary hypothesis tests to decide membership within a non-convex set and self-tests of non-extreme operations are not possible dropping IID. The first result implies that non-convex sets of inequalities for the analysis of quantum networks are not sound in absence of IID conditions. This means that in a causal network without inputs, a source of shared randomness allows all possible correlations to be achieved with just classical resources. The second result shows that, in the non-IID regime it is only possible to self-test *extreme* states and POVMs. In this context "extreme" means pure states and measurements which cannot be expressed as (non-trivial) convex combinations of other operators.

- **Eavesdropper detection:** Regarding the eavesdropper detection, there are two ways that the string S_{test} (Section 4.1.2) could be used: as a selector, or grouping bits together. Imagine that $S_{test} = 011010100110$. In the first scenario, of the $n_{test} + n_{key}$ rounds those associated to a 1 will be used for testing (second, third, fifth, ...), while those associated to a 0 will be used for key generation (first, fourth, ...). In the second scenario, we could rewrite S_{test} as $\{011\}\{010\}\{100\}\{110\}$, and convert each number in the brackets to decimal. In this way, of the first eight (2^3) rounds of $n_{\text{test}} + n_{key}$ the sixth (011) would be for testing, and the others would be for key generation. Of the next eight rounds, the fourth (010) would be for detection et cetera. The selector scenario allows for a higher number of testing rounds, and thus a shorter key, while the grouping scenario allows for sparser testing and a longer key, the needs and resources of the users dictate which is more convenient.

Acknowledgements

This project was done with the support of the University of Bologna and of ICFO

Firstly I want to thank Professor Piroli for the attention and care he put in supervising me, following my work before, during and after my stay in Barcelona. I want to equally thank Professor Acín for taking me in for this project in his research group, giving me an opportunity unlike any other.

This work is the reflection of the help and support I received from so many people.

I am enormously grateful to Raja and Tamás, you did an incredible job mentoring me and your guidance was essential to the development of this project. Your trust created an environment where I could grow and learn, and your care helped me feel safe in a new place doing a new thing.
I want to thank Raffaele as well, with which I shared much of the work and who helped me along the way.

I want to express my gratitude to all the people I met at ICFO, who welcomed me with open arms, and to those I met outside the workplace. You've made my stay much lighter and my leaving much heavier.

A very big thank you goes to all the friends I made in Bologna and to my incredible housemates: sharing all this with you has been wonderful and hopefully it won't end here. If I felt at home here is because of your affection and closeness.
A very big thank you goes to all the friends from Pistoia as well, growing up with you was the best possible way to spend all these years.

I am extremely grateful to my family, without your support I wouldn't have arrived this far. There were, there are and there most surely will be hard times, but if I am the person I'm proud to be, it is also thanks to you. This success is yours too.

My deepest gratitude goes to Giovanna. Your support and your love go beyond what words can express, thank you for being by my side anytime and in any condition. Every day I spent with you made me love you more, and every day I spent away from you made me appreciate what we built.

Bibliography

[1] Acin, A., Brunner, N., Gisin, N., Massar, S., Pironio, S., and Scarani, V. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters 98*, 23 (June 2007).

[2] Acin, A., Caramelli, L., D'Avino, R., Krivachy, T., Pousa, R. G., Senno, G., and Yehia, R. QKD protocol without inputs [Manuscript in preparation] (n.d.).

[3] Arqand, A., and Tan, E. Y.-Z. Marginal-constrained entropy accumulation theorem. *arXiv preprint* (July 2025). arXiv:2502.02563 [quant-ph].

[4] Barrett, J., Linden, N., Massar, S., Pironio, S., Popescu, S., and Roberts, D. Non-local correlations as an information theoretic resource. *Physical Review A 71*, 2 (Feb. 2005). arXiv:quant-ph/0404097.

[5] Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika 1*, 3 (1964).

[6] Bennett, C. H., and Brassard, G. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science 560* (Dec. 2014), 7–11.

[7] Bennett, C. H., Brassard, G., and Mermin, N. D. Quantum cryptography without Bell's theorem. *Physical Review Letters 68*, 5 (Feb. 1992).

[8] Bennett, C. H., DiVincenzo, D. P., Fuchs, C. A., Mor, T., Rains, E., Shor, P. W., Smolin, J. A., and Wootters, W. K. Quantum Nonlocality without Entanglement. *Physical Review A 59*, 2 (Feb. 1999). arXiv:quant-ph/9804053.

[9] Boreiri, S., Krivachy, T., Sekatski, P., Girardin, A., and Brunner, N. Topologically Robust Quantum Network Nonlocality. *Physical Review Letters 134*, 1 (Jan. 2025). arXiv:2406.09510 [quant-ph].

[10] Boreiri, S., Ulu, B., Brunner, N., and Sekatski, P. Noise-robust proofs of quantum network nonlocality. *Quantum 9* (Aug. 2025). arXiv:2311.02182 [quant-ph].

[11] Canetti, R. Universally composable security: a new paradigm for cryptographic protocols. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science* (Oct. 2001), IEEE.

[12] Donohue, J. M., and Wolfe, E. Identifying Nonconvexity in the Sets of Limited-Dimension Quantum Correlations. *Physical Review A 92*, 6 (Dec. 2015). arXiv:1506.01119 [quant-ph].

[13] Duncan, T. L., and Semura, J. S. Information Loss as a Foundational Principle for the Second Law of Thermodynamics. *Foundations of Physics 37*, 12 (Nov. 2007). arXiv:cond-mat/0703235.

[14] Dupuis, F., and Fawzi, O. Entropy Accumulation With Improved Second-Order Term. *IEEE Transactions on Information Theory 65*, 11 (Nov. 2019).

[15] Dupuis, F., Fawzi, O., and Renner, R. Entropy accumulation. *Communications in Mathematical Physics 379*, 3 (Nov. 2020). arXiv:1607.01796 [quant-ph].

[16] Einstein, A., Rosen, N., and Podolsky, B. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review 47*, 10 (1935).

[17] Ekert, A. K. Quantum cryptography based on Bell's theorem. *Physical Review Letters 67*, 6 (Aug. 1991).

[18] Fritz, T. Beyond Bell's theorem: correlation scenarios. *New Journal of Physics 14*, 10 (Oct. 2012).

[19] Galvão, E. F., and Brod, D. J. Quantum and classical bounds for two-state overlaps. *Physical Review A 101*, 6 (June 2020). arXiv:1902.11039 [quant-ph].

[20] Gisin, N., and Popescu, S. Spin flips and quantum information for antiparallel spins. *Physical Review Letters 83*, 2 (July 1999).

[21] Gisin, N., Ribordy, G., Tittel, W., and Zbinden, H. Quantum Cryptography. *Reviews of modern physics 74*, 1 (Mar. 2002).

[22] Henson, J., Lal, R., and Pusey, M. F. Theory-independent limits on correlations from generalised Bayesian networks. *New Journal of Physics 16*, 11 (Nov. 2014).

[23] Jaynes, E. T. Information Theory and Statistical Mechanics. *Physical Review 106*, 4 (1957).

[24] Jaynes, E. T. Information Theory and Statistical Mechanics. II. *Physical Review 108*, 2 (1957).

[25] Koenig, R., Renner, R., and Schaffner, C. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory 55*, 9 (Sept. 2009). arXiv:0807.1338 [quant-ph].

[26] Le, T. P., Meroni, C., Sturmfels, B., Werner, R. F., and Ziegler, T. Quantum Correlations in the Minimal Scenario. *Quantum 7* (Mar. 2023). arXiv:2111.06270 [quant-ph].

[27] Metger, T., Fawzi, O., Sutter, D., and Renner, R. Generalised entropy accumulation. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (Oct. 2022). arXiv:2203.04989 [quant-ph].

[28] Murta, G., Grasselli, F., Kampermann, H., and Bruss, D. Quantum Conference Key Agreement: A Review. *Advanced Quantum Technologies 3*, 11 (Nov. 2020). arXiv:2003.10186 [quant-ph].

[29] Navascues, M., and Wolfe, E. The Inflation Technique Completely Solves the Causal Compatibility Problem. *Journal of Causal Inference 8*, 1 (Sept. 2020).

[30] Nielsen, M. A., and Chuang, I. L. *Quantum computation and quantum information*, 10th anniversary edition ed. Cambridge university press, Cambridge, 2010.

[31] Park, J. L. The concept of transition in quantum mechanics. *Foundations of Physics 1*, 1 (1970).

[32] Pirandola, S., Andersen, U. L., Banchi, L., Berta, M., Bunandar, D., Colbeck, R., Englund, D., Gehring, T., Lupo, C., Ottaviani, C., Pereira, J., Razavi, M., Shaari, J. S., Tomamichel, M., Usenko, V. C., Vallone, G., Villoresi, P., and Wallden, P. Advances in Quantum Cryptography. *Advances in Optics and Photonics 12*, 4 (Dec. 2020). arXiv:1906.01645 [quant-ph].

[33] Preskill, J. Lecture Notes for Ph219/CS219: Quantum Information and Computation.

[34] Renner, R. *Security of Quantum Key Distribution*. PhD thesis, Jan. 2006. arXiv:quant-ph/0512258.

[35] Renou, M.-O., and Beigi, S. Network Nonlocality via Rigidity of Token-Counting and Color-Matching. *Physical Review A 105*, 2 (Feb. 2022).

[36] Renou, M.-O., Bäumer, E., Boreiri, S., Brunner, N., Gisin, N., and Beigi, S. Genuine quantum nonlocality in the triangle network. *Physical Review Letters 123*, 14 (Sept. 2019). arXiv:1905.04902 [quant-ph].

[37] Ribeiro, J., Murta, G., and Wehner, S. Fully device independent Conference Key Agreement. *Physical Review A 97*, 2 (Feb. 2018). arXiv:1708.00798 [quant-ph].

[38] Rivest, R. L., Shamir, A., and Adleman, L. A method for obtaining digital signatures and public-key cryptosystems. *Communication of the ACM 21*, 2 (Feb. 1978).

[39] Scarani, V. *Bell Nonlocality*, 1 ed. Oxford University PressOxford, Aug. 2019.

[40] Scarani, v., Acin, A., Ribordy, G., and Gisin, N. Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations. *Physical Review Letters 92*, 5 (Feb. 2004).

[41] Sekatski, P., Boreiri, S., and Brunner, N. Partial self-testing and randomness certification in the triangle network. *Physical Review Letters 131*, 10 (Sept. 2023). arXiv:2209.09921 [quant-ph].

[42] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal 27*, 3 (July 1948).

[43] Shor, P. W. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Journal on Computing 26*, 5 (Sept. 1997).

[44] Supic, I., Bancal, J.-D., Cai, Y., and Brunner, N. Genuine network quantum nonlocality and self-testing. *Physical Review A 105*, 2 (Feb. 2022).

[45] Supic, I., and Bowles, J. Self-testing of quantum systems: a review. *Quantum 4* (Sept. 2020).

[46] Tavakoli, A., Gisin, N., and Branciard, C. Bilocal Bell inequalities violated by the quantum Elegant Joint Measurement. *Physical Review Letters 126*, 22 (June 2021).

[47] Tavakoli, A., Pozas-Kerstjens, A., Luo, M.-X., and Renou, M.-O. Bell nonlocality in networks. *Reports on Progress in Physics 85*, 5 (May 2022). arXiv:2104.10700 [quant-ph].

[48] Tomamichel, M. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETHz, July 2012. arXiv:1203.2142 [quant-ph].

[49] Tomamichel, M., Colbeck, R., and Renner, R. A Fully Quantum Asymptotic Equipartition Property. *IEEE Transactions on Information Theory 55*, 12 (Dec. 2009). arXiv:0811.1221 [quant-ph].

[50] Tomamichel, M., Schaffner, C., Smith, A., and Renner, R. Leftover Hashing Against Quantum Side Information. *IEEE Transactions on Information Theory 57*, 8 (Aug. 2011). arXiv:1002.2436 [quant-ph].

[51] Watrous, J. *The Theory of Quantum Information*, 1 ed. Cambridge University Press, Apr. 2018.

[52] Watrous, J. Lecture notes on Advanced topics in quantum information theory, 2020.

[53] Weilenmann, M., Budroni, C., and Navascues, M. Memory attacks in network nonlocality and self-testing. *Quantum 9* (May 2025). arXiv:2402.12318 [quant-ph].

[54] Wiesner, S. Conjugate coding. *ACM SIGACT News 15*, 1 (Jan. 1983).

[55] Wolf, R. *Quantum key distribution : an introduction with exercises*, 1st ed. ed. Springer, Cham, Switzerland, 2021.

[56] Wolfe, E., Spekkens, R. W., and Fritz, T. The Inflation Technique for Causal Inference with Latent Variables. *Journal of Causal Inference 7*, 2 (Sept. 2019).

[57] Yehia, R., Diamanti, E., and Kerenidis, I. Composable Security for Multipartite Entanglement Verification. *Physical Review A 103*, 5 (May 2021).