Projecte de Final de Carrera

Extracció de secret a partir correlacions no-signaling (Secrecy extraction from non signaling correlations)

Escola Tècnica Superior d'Enginyeria de Telecomunicacions de Barcelona

Estudis: Enginyeria de Telecomunicació
Autor: Sergi Pino Paulino
Director: Antonio Acín dal Maschio

Juny 2016

Dedicado a la memoria de A.P.R.

Colaboracions

El present treball fou realitzat durant el quatrimestre de primavera 2006 a les instalacions de l'Institut de Ciències Fotòniques, al Parc Mediterrani de la Tecnologia, Castelldefels.

The first principle is that you must not fool yourself and you are the easiest person to fool.

Richard P. Feynman

Recognize reality even if you don't like it, specially if you don't like it.

Charles T. Munger

Prefaci i Agraïments

Aquest Projecte de Final de Carrera és el resultat d'un apropament de l'autor al camp de la informació quàntica, en particular la criptografia quàntica -o més ben dit la distribució quàntica de claus- un camp molt interdisciplinar en el que conflueixen aspectes de la mecànica quàntica, la teoria de la informació i la matemàtica aplicada (probabilitat, optimització,...).
Part dels resultats que aquí es presenten apareixen en un article en el que l'autor va col·laborar al seu moment [18] i es va publicar en el número 74 de la Physical Review A.
Aquest PFC es va realitzar durant el quatrimestre de primavera del curs 2005-2006 a l'ICFO, Institut de Ciències Fotòniques, al Parc Mediterrani de la Tecnologia.
Vull agraïr al meu director, en Toni Acín, per haver-me permès participar en aquest projecte i per les discussions tant enriquidores que vam tenir sobre el tema al seu moment. També vull agraïr-li que hagi tingut la generositat de permetre'm finalitzar el que vaig deixar sense finalitzar tant de temps després. Vull agrair als meus pares pel seu suport incondicional durant tota la meva vida, el seu recolzament en els moments difícils i de dubtes i per haver estat sempre al meu costat. I també a la archi pel suport i la paciència en la fase final d'aquesta història. He trigat 10 anys en redactar aquest document. Si no hagués tingut la obligació deguda al canvi de plans d'estudis, podria haver trigat 10 anys més.

El resultat no és el que m'hauria agradat al seu moment, però la demora no ha estat deguda a un excès de perfeccionisme. Diguem, sense entrar en justificacions adicionals, que la vida no m'ha donat ni el temps ni la força de voluntat necessaris per enfrontar-me als records que m'assaltaven al treballar en aquesta memòria.

Voldria, abans d'iniciar la memòria del projecte, respondre a dues preguntes de manera breu:

- Perquè em vaig decidir a fer aquest PFC al seu moment i que em va aportar?
  En el moment d'estar a punt de finalitzar la carrera d' Enginyeria de Telecomunicacions i també havent estudiat Matemàtiques, tenia interés en aprendre sobre algun tema nou i evaluar si la meva manera de ser aleshores era apta per la recerca (la resposta a aquesta última pregunta va ser no). Adicionalment, com no havia trobat fins a aquell moment un àmbit del coneixement adquirit a la carrera al que dedicar-me (una passió, amb la concepció errònia de que la plenitut personal provenia de trobar-la), volia interactuar amb temàtiques diferents. El tema que vaig escollir era intelectualment atractiu i era prou diferent al que havia fet fins al moment com per a tenir curiositat. A més la mecànica quàntica sempre m'havia semblat un tema molt atractiu, per les lectures casuals que n'havia fet i les petites incursions que havia realitzat. Sota el punt de vista de la realització d'un projecte de final de carrera en Enginyeria de Telecomunicacions, la confluència d'elements de teoria de la informació, física i matemàtica aplicada em semblava adecuada. No negaré que tenia dubtes sobre si la temàtica encaixaria en un PFC de "telecos", però amb la perspectiva del temps transcorregut, sembla cada cop més clar que la informació quàntica és un àmbit del coneixement tant de la física com del món del processament de la informació.
  Poder haver estat físicament en un centre de recerca punter a nivell mundial va ser una experiència molt enriquidora.

- Que vaig aportar amb la realització d'aquest PFC?

  La meva participació va consistir en, a part de l'enteniment dels protocols i principis bàsics descrits en les següents pàgines, en analitzar les fites de $KeyRate$ proposades i obtenir els resultats referents a la (no) nulitat de la informació intrínseca, tant en els cas $d = 2$ com $d = 3$. L'enteniment del formalisme que es fa servir per a desenvolupar el protocol de distribució de

claus va ser una tasca àrdua, no tant per la complexitat de les eines sino per la familiarització amb conceptes que són, parafrasejant a Niels Bohr, difícils d'entendre.

La fita final del projecte comú en el que vaig participar va suposar la publicació de l'article [18], en el que vaig participar.

Resum del projecte

La criptografia quàntica és el camp de la teoria de la informació quàntica que permet fer tasques criptogràfiques, és a dir de generació de secret entre parts, a partir dels principis de la mecànica quàntica. Si bé els principis i el fonament teòric que permeten distribuir correlacions amb secret a partir de l'entrellaçament són vàlids, s'ha d'assegurar que no es perd el secret per qualsevol desviament del cas ideal i per això molta recerca teòrica s'ha dedicat a derivar fites riguroses per la seguretat de la criptografia quàntica. Les proves de seguretat usuals suposen que els participants en l'intercanvi de claus, Alice i Bob, tenen un coneixement perfecte i control sobre els seus sistemes i aparells quàntics. Per exemple, han d'estar segurs que els seus bits s'han codificat en qubits i no en sistemes de més dimensions que permetin que un atacant pugui extreure'n informació.

Davant d'aquest problemàtica, l'estudi de la no-localitat intrínseca de la mecànica quàntica com a recurs d'informació més enllà de la pròpia mecànica quàntica, va proporcionar una possible solució. Les correlacions que apareixen en el fenòmen físic de l'entanglement violen el que es coneixen com a desigualtats de Bell, que limiten les correlaciones que es poden obtenir via models de variables locals (o equivalentment, estratègies deterministes). L'estudi de l'estructura de correlacions entre les parts comporta l'estudi de l'anomenat politop non-signaling que està format per cares definides per les desigualtats de Bell (el politop local) i punts extremals no-locals (que suposen una violació maximal de les desigualtats de Bell).

A partir de la comprensión de la estructura del politopo, se presenta un protocolo de distribución de claves basado exclusivamente en el grado de violación de las desigualdades de Bell por parte de las correlaciones entre las partes ante un atacantes limitado exclusivamente por la restricción no-signaling (no es posible transmitir información a mayor velocidad que la de la luz).

En aquest projecte, l'autor va poder participar en l'estudi i generalització d'aquest protocol de distribució de claus sota assumpció de correlacions non-signaling, estudi de l'estructura de les correlacions i obtenció de fites per a la generació de clau i fites teòriques de la informació intrínseca pel protocol amb número de possibles outputs $d = 2$ i $d = 3$.

Amb posterioritat als resultats aquí presentats, el grup de Toni Acín i col·laboradors ha extés aquesta línia de recerca per a provar la seguretat de protocols quàntics de distribució de claus independents del dispositiu (el que s'anomena Device Independent Quantum Key Distribution - DIQKD-) davant d'atacs més generals.

La resta del document estarà redactat en anglès per comoditat en la terminologia emprada.

Resumen del proyecto

La criptografía cuántica es el campo de la teoría de la información cuántica que permite realizar tareas criptográficas, es decir de generación de secreto entre partes, a partir de los principios de la mecánica cuántica.

Si bien los principios y fundamento teórico que permiten distribuir correlaciones con secreto a partir de correlaciones son válidos, se debe asegurar que no se pierde el secreto por cualquier desviación del caso ideal y por este motivo, esfuerzos de investigación teórica se han dedicado a derivar cotas rigurosas para la seguridad de la criptografía cuántica.

Las pruebas de seguridad habituales suponen que los participantes en el intercambio de claves, Alice y Bob, poseen un conocimiento perfecto y control total sobre sus sistemas y aparatos cuánticos. Por ejemplo, deben estar seguros que sus bits se han codificado en qubits y no en sistemas de un número de dimensiones más elevado que permita que un atacante pueda extraer información de éstas.

Ante esta problemática, el estudio de la no-localidad intrínseca de la mecánica cuántica como recurso de información más allá de la propia mecánica cuántica, proporciona una posible solución. Las correlaciones que aparecen en el fenómeno físico del entanglement violan lo que se conocen como desigualdades de Bell, que limitan las correlaciones que se pueden obtener por medio de modelos de variables locales (o equivalentemente, estrategias deterministas). El estudio de la estructura de correlaciones entre las partes comporta el estudio del denominado politopo non-signaling que está formado por caras definidas por las desigualdades de Bell (el politopo local) y puntos extremales no-locales (que conllevan una violación máximal de las desigualdades de Bell).

A partir de l'enteniment de l'estructura del politop, es presenta un protocol de distribució de claus que es basa exclusivament en el grau de violació de desigualtats de Bell de les correlacions entre les parts davant d'un atacant només limitat per la restricció no-signaling (no es pot transmetre informació a major velocitat que la de la llum).

En este proyecto, el autor pudo participar en el estudio y generalización de este protocolo de distribución de claves bajo asunción de correlaciones non-signaling, estudio de la estructura de las correlaciones y obtención de cotas para la generación de claves y cotas teóricas de la información intrínseca para el protocolo con un número posible de outputs $d = 2$ y $d = 3$.

Con posterioridad a los resultados aquí presentados, el grupo de Toni Acín y colaboradores han extendido esta línea de investigación para probar la seguridad de protocolos cuánticos de distribución de claves independientes del dispositivo (lo que se denomina Device Independent Quantum Key Distribution -DIQKD-) ante ataques más generales.

El resto del documento está redactado en inglés por comodidad con la terminología utilizada.

Abstract

Quantum cryptography studies tasks in Quantum Information Theory for cryptographic purposes, that is, for secret generation between parts from the principles of quantum mechanics.

The fact itself, that quantum correlations can be used to distribute secrecy, is safe. But of course, one must verify that secrecy is not immediately spoiled by any small departure from this ideal case; this is why much theoretical research has been devoted to the derivation of rigorous bounds for the security of quantum cryptography

Usual security proofs contemplates that the participants in the key distribution, Alice and Bob, have perfect knowledge and total control over their quantum systems and devices. For example, one assumes that the logical bits are encoded in quantum systems whose dimension is under perfect control (generally, qubits). Additional dimensions of the quantum system could be exploited by an attacker to extract information (side channels).

The study of intrinsic non-locality of quantum mechanics as an information resource provides a possible solution to this problem. The correlations obtained from the physical phenomenon of entanglement violate Bell inequalities, that limit the correlations that can be obtained with local variables models (or equivalently, deterministic strategies). The study of the structure of correlations between the parties is the study of the non-signaling polytope which is limited by faces defined by Bell inequalities (local polytope) and extremal non-local points (representing a maximal violation of Bell inequalities).

From the study of the polytope's structure, a key distribution protocol is developed which is based solely in the degree of violation of Bell inequalities achieved by the correlations between the parties. Moreover, the attacker is limited only by the non-signaling restriction - there is no instant transmission of information-.

In this project, the author could participate in the study and generalization of said key distribution protocol based in non-signaling correlations, the study of the correlations' structure and the derivation of secrecy generation bounds and theoretical bounds for intrinsic information, for the protocol with number of outputs $d = 2$ and $d = 3$.

Afterwards the results here presented, Toni Acín's group and collaborators extended this research path to prove the security of quantum key distribution protocols independently of the device(what is called Device Independent Quantum Key Distribution -DIQKD-) in front of more general attacks.

The rest of the document will be written in English.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Quantum information theory combines two of the most important scientific discoveries of the 20th century: Information theory initially developed by Claude Shannon and Quantum Mechanics. Quantum Information Theory addresses two issues under the author's point of view:

- It applies the physical theory of quantum mechanics to communication and information processing in which the theory is relevant (using entangled states or where following Moore's law, quantum mechanics matters as a consequence of the reduction of size).
- By studying these applications, it is a very useful tool to understand physical reality and interpret the counter-intuitive results of quantum physics.

If one had to enumerate the most quantum features, despite the name of the theory, these would be superposition and entanglement. Both concepts are widely applied in different instances of quantum information theory. On the one hand, in quantum computation, superposition of states allow one to perform computing tasks unimaginable in classical terms. On the other hand, entanglement and non-locality are the elements shown to provide a mean to distribute correlations at a distance, whose secrecy can be guaranteed by the laws of physics, without any assumption on the computational power of the eavesdropper. The employment of quantum physics as a tool to obtain secrecy is the largely studied field of quantum cryptography (or more accurately, quantum key distribution, QKD). This the most mature development of quantum information science; quantum cryptography is already a commercial product and a tool that is employed by companies and institutions to achieve secrecy in their communication.

The secrecy obtained from these schemes is guaranteed by the physical principles of nature and the laws of nature are not going to change in the near future... But, as Yogi Berra said "In theory there is no difference between theory and practice. In practice there is.": the implementation of the ideal case can suppose a small departure that can turn the whole key distribution scheme insecure, not because its principles are wrong, but because the scheme can be exploitable by a malicious attacker. A lot of theoretical research has been devoted to derive rigorous bounds for the security of quantum cryptography, establishing the conditions under which a secret key can be extracted. But these security proofs can be just theoretical, and once the cryptographic schemes are put into "practice", the implementation can have different assumptions from the ideal case.

In particular, an assumption present in theoretical proofs unnoticed before the works presented in [2, 5] was that the encoding of the bits in the quantum system is done having under control the dimension of such systems. Removing this assumption has interest per se, but also side channels in a device can be a serious problem from the point of view of the implementation. For example, the designers of a device have to be careful that when they encode (say) polarization, they encode only polarization, and that the device does not change the spectral line, or the spatial mode, or the temporal mode of the photon as well. And these features could be exploitable maliciously. The hope lies in the realization of Device Independent Quantum Key distribution schemes.

In the work described in this memoir, an approach that was a first step in that direction is presented. Employing recent (at the time) developments in the understanding of non-locality from an information theoretic point of view and as an information resource, a key distribution protocol was developed in [2]. The defined protocols, both for the case of bits and for generic d-dimensional outcomes, are extensions of this work. As in every typical quantum key distribution scheme, the

trusted parties Alice and Bob aim to generate a secret key while the eavesdropper, Eve intends to obtain the secret key.

The idea of the protocols here studied resides in guarantee security and the possibility of key extraction by the structure of the Alice-Bob correlations, under the no-signaling[1] condition. A protocol or scheme to distribute secrecy is based on the idea that if the correlations cannot be produced by pre-established or local deterministic strategies, then Eve has poor knowledge of Alice's and Bob's symbols. Additionally, if the protocol reduce the attacker information in the cases where shared randomness is employed, the secrecy extraction can be improved.

Moreover, there is an extra assumption: Eve has total control over the devices but not over the choice of measurements performed. It is Eve who defines the correlations between Alice and Bob. A pictorial description of the no-signalling assumption in the context of secrecy extraction is necessary. This is shown in 1.1 The dark grey boxes in Alice's and Bob's laboratories are the devices provided by Eve. In a first step, the laboratories are open for the signal that correlates them (grey spheres). The arrows on the channel indicate that it is not important whether this signal comes from outside, or is emitted by Alice's device to Bob's, or viceversa: in any case, it is under Eve's control. What is important, is that the inputs (x, y) have not been chosen yet. In a second step, the laboratories are absolutely closed: no leakage of information about the inputs (x, y) or the outputs (a, b) is allowed. In a third step (not shown), Alice and Bob can carry out the usual procedures of error correction and privacy amplification by communicating on an authenticated channel.

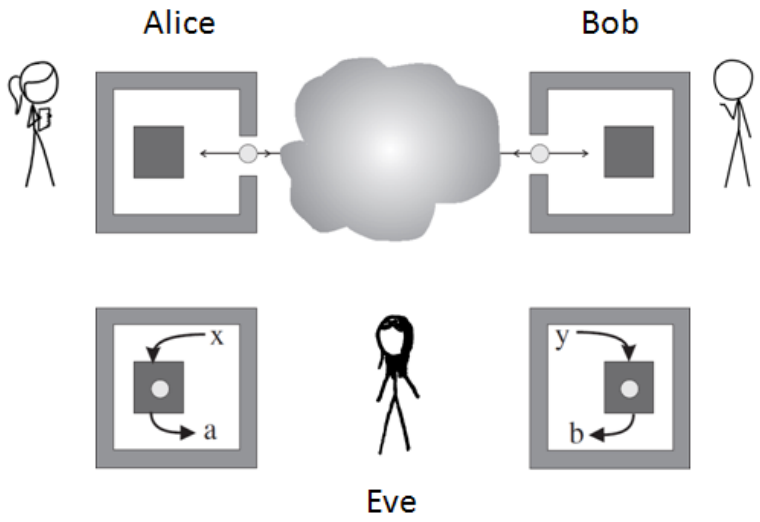


Figure 1.1: Scheme of the protocol

The present study assumes, on the one hand that the eavesdropper Eve performs only individual[2] attacks, on the other hand that Eve can distribute any correlation compatible with the no-signaling condition (in this sense her power is greater than what quantum physics allows). Under these assumptions, we prove that the protocols defined here allow extracting secrecy from noisy correlations, when these correlations violate a Bell-type inequality by a sufficiently large amount. The region, in which secrecy extraction is possible, extends within the region of correlations achievable by measurements on entangled quantum states.

The following pages are the description of the necessary concepts and tools to introduce the cryptographic protocols under study and the study of these protocols.

[1]We will define formally the no-signaling condition in the text, but it can be described as "faster than light communication is impossible" or the probability distribution of the outcomes of each one of the parties just depends on his choice of measurement

[2]Eve sends independent identically distributed signals and she tries to guess each bit of the raw key without performing correlations between instances. In chapter 5, the different types of attack will be defined.

Memoir structure

The structure of the present document is the following:

- A first chapter with an introduction to notions from Quantum Information Theory. The objective is to introduce the concept of entanglement and Bell inequalities.
- A second chapter describing the tools from an information-theoretic point of view necessary to generate a symmetric key between two parties. The concepts introduced will be privacy amplification, advantage distillation and intrinsic information.
- A third chapter describing;
  - the framework to study non-locality independent of the Quantum Mechanics formalism.
  - the structure of the correlations in this formalism and the Bell inequalities involved in $d = 2$ and larger-dimensional outcomes.
  - Results equivalent to the ones of quantum information theory in the non-signaling setting.
- A fourth chapter describing the protocol to distribute a secret key against an eavesdropper not limited by quantum physics, employing individual attacking strategies based on the violation of Bell inequalities and some classical information processing. The following results will be detailed:
  - Description of the protocol and structure of the correlations involved.
  - For the binary case: analysis of results derived from symmetries of these correlations, derivation of bounds to generate secret key in different settings and analysis of the intrinsic information between the parties.
  - For larger-dimensional outcomes: structure of the correlations involved, bounds to extract secret keys and analysis of the intrinsic information.

The results of the work here presented was published in the number 74 of Physical Review A ([18])

Chapter 2

Quantum Information and Quantum Cryptography

This chapter describes briefly the foundations of quantum mechanics and quantum information theory necessary for the rest of the document. Firstly, we give a summary of the mathematical techniques of the quantum formalism which are the rules common to any Quantum Information application. Then we present the quantum bit or qubit. We describe the phenomenon of entanglement and some of its consequences from the point of view of quantum information applications. Then we present several examples of quantum communication protocols and introduce briefly Quantum Cryptography.

Although, the protocol described is a quantum based key distribution protocol, the mathematics employed to describe it are not the mathematics of quantum mechanics, but instead, as we will see in chapter 4, a more general way of describing the phenomena experienced in entanglement.

We introduce these concepts here because they are the entourage in which the protocol analyzed resides and the implementation of said protocol would be done with quantum states. Additionally, it's necessary to introduce the concept of entanglement and Bell inequalities as the test to identify non-local correlations.

2.1 Quantum mechanics for Quantum information theory

The scope of this first section is to introduce the basic postulates and mathematical formalism of Quantum Mechanics. Most of the formalism presented here can be found in chapter 2 of [11].
Briefly stated, all quantum machinery is linear algebra on a complex Hilbert space.

Postulate 1 *Any isolated physical system has associated a complex Hilbert space $\mathcal{H}$ (vector space with inner product). The description of this system is given by a normalized vector in this space, the state vector which is an element of $\mathbb{P}\mathcal{H}$, i.e. a unit vector.*

Quantum Mechanics does not tell what the state space for a determined physical system nor does it tell us what the state vector of the system is, it provides a formalism that comprises any physical state. The spaces employed in Quantum Information Theory are finite dimensional Hilbert spaces, $\mathbb{C}^2$ to work with *qubits* and $\mathbb{C}^d$ to work with *qudits*.
The standard notation employed to explain the quantum machinery is the Dirac Notation:

- $|\psi\rangle$ stands for a column vector, also known as a *ket*
- $\langle\psi|$ stands for the dual vector to $|\psi\rangle$, also known as *bra*
- $\langle\phi|\psi\rangle$is the scalar product between $|\psi\rangle$ and $|\phi\rangle$

The norm of a vector $|\psi\rangle$ is $\| \psi \| = \sqrt{\langle\psi|\psi\rangle}$.
Following from the first postulate, if $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^d$ are two possible states of the system, then any possible linear combination of these two vectors $|\psi\rangle = \alpha|\psi_1\rangle + \beta|\psi_2\rangle$, where α and β are complex numbers and such that $\| \psi \| = 1$, also represents a valid description of the system. This is known

as the superposition principle and $|\psi\rangle$ is usually called a coherent superposition of $|\psi_1\rangle$ and $|\psi_2\rangle$. Therefore, all the information on an isolated physical system can be specified through a vector in a Hilbert space.

The evolution of quantum systems is prescribed by the postulate:

Postulate 2 *The evolution of a closed quantum system is described by a unitary transformation, that is an operator U such that $U^\dagger U = 1$. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system at the time t_2 by a unitary operator U which depends only on the times t_1 and t_2*

$$|\psi'\rangle = U|\psi\rangle$$

The next postulate of quantum mechanics is the one concerning quantum measurement that describes the interaction of an experimentalist with a quantum system she observes.

Postulate 3 *Quantum measurements are described by a collection $\{M_m\}$ of measurement operators. These are operators acting on the state space being measured. The index* m *refers to the measurement outcome that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ just before the measurement then the probability that result* m *occurs is*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$$

and the state after the measurement is

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_n|\psi\rangle}}$$

The measurement operators satisfy the completeness equation,

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

The completeness equation expresses the fact that probabilities sum to one:

$$1 = \sum_m p(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle \ \forall|\psi\rangle$$

Consider a measurement, in which different positions of the pointer can be associated with d different states $|\phi_1\rangle, \ldots |\phi_1\rangle$. Among the features of an ideal measurement, one tends to request that the device identifies the state correctly.

The first postulate states that pure states are described by one-dimensional spaces (every vector $c|\psi\rangle$ differing by a constant c represents the same state). With this construction, after a measurement of the system, it may be found to possess a property that it did not possess with certainty before the measurement. Given a state $|\psi_1\rangle$, there is a non-zero probability that a measurement finds the system in a different state $|\psi_2\rangle$. The probability is given by:

$$P(\psi_2 \mid \psi_1) = \mathrm{tr}(P_1 P_2) = \mid \langle\psi_1|\psi_2\rangle \mid^2$$

This is called the Born's rule for probabilities.

The next postulate deals with the description of composite systems, quantum systems made up of two or more distinct physical system.

Postulate 4 *the state space of a composite system is the tensor product of the state space of the component physical systems. Moreover, if we have systems numbered* 1 *through n, and system i is prepared in the system $|\psi_i\rangle$, then the state of the total system is the tensor product $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$*

This postulate can also be expressed in a more simple and convenient way for our main purpose: Consider two physical systems A and B described by the corresponding Hilbert spaces, $\mathcal{H}_A$ and

$\mathcal{H}_B$. The Hilbert space associated to the global system AB, denoted by $\mathcal{H}$, consists on the tensor product of the two local spaces, $\mathcal{H}_A \otimes \mathcal{H}_B$. Something to note of enormous importance, is that when this postulate and the superposition principle are combined the concept of entanglement[1] arises. Indeed, consider two possible states of the global system $\mathcal{H}_A \otimes \mathcal{H}_B$, $|\psi_A\rangle \otimes |\psi_B\rangle$ and $|\phi_A\rangle \otimes |\phi_B\rangle$. then $|\Psi\rangle = \alpha|\psi_A\rangle \otimes |\psi_B\rangle + \beta|\phi_A\rangle \otimes |\phi_B\rangle$ is also a possible state of the composite system AB. However, this state in general cannot be written as the tensor product of two vectors on each local space, that is $|\Psi\rangle \neq |\varphi_A\varphi_B\rangle$. The state $|\Psi\rangle$ is then said to be entangled. Although the global state of the system can be described by a vector of the composite system, it is impossible to associate a vector in $\mathcal{H}_A$ and $\mathcal{H}_B$. This is related to the fact that system A is correlated to system B. In following sections, more properties will be described about this principle as entanglement states play a crucial role in quantum information and are relevant for the cryptographic protocol that will be described at the end of this memoir.
Examples of entangled systems of two qubits are the Bell states. The state $|\Psi^-\rangle$ is also called the *singlet*.

$$\begin{aligned} |\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\ |\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ |\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \end{aligned} \tag{2.1}$$

Consider the case where the preparation of a system is imperfect, the system can be in state $|\psi_1\rangle$ with probability p_1 and in state ψ_2 with probability p_2, and so on. In this case, as in the case of entangled systems, there is a lack of knowledge about the state of the system. Since the information on the state of the system is not complete, the state cannot be pure. In order to take into account this lack of information, the so-called mixed states have to be introduced.
The mathematical description of a system that can be in state $|\psi_i\rangle \in \mathcal{H}$ with probability p_i, where $i = 1, \ldots, N$ and N is arbitrary, is given by an operator called the density operator or density matrix:

$$\rho = \sum_{i=1}^{N} p_i |\psi_i\rangle\langle\psi_i|$$

Note that $\mathrm{tr}(\rho) = \sum_i p_i \,\mathrm{tr}(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i \langle\psi_i|\psi_i\rangle = \sum_i p_i = 1$. If the state is pure, $|\psi\rangle$, there is only one no-vanishing probability, $p_1 = 1$, and $\rho = |\psi\rangle\langle\psi|$. Actually a state ρ is pure if and only if $\mathrm{tr}\,\rho^2 = 1$. Once a mixed state is defined, one can specify how an open quantum system evolves or how to describe a noisy evolution. In this case, a quantum system in a pure state may lose its purity and become mixed. The general formalism is given by the so-called trace preserving maps. Indeed, any evolution of an initial state, possibly mixed, ρ_i, into a state ρ_f, can be represented as:

$$\rho_f = \sum_k A_k \rho_i A_k^\dagger$$

where A_k are operators such that $\sum A_k A_k^\dagger = \mathbb{1}$, the identity matrix.
Combining the mixed-state formalism with postulate 3, one can specify a general evolution, consisting of a sequence of measurements and possibly noisy evolutions, through a set of operators A_k^i describing a map transformation from any state ρ into:

$$\rho_i = \frac{\sum_k A_k^i \rho A_k^{i\dagger}}{\mathrm{tr}(\sum_k A_k \rho A_k^\dagger)}$$

with probability $p_i = \mathrm{tr}(\sum_k A_k \rho A_k^\dagger)$. The standard von Neumann measurement formalism is described by the previous one. A von Neumann, or projective measurement in a d-dimensional system has d possible outcomes and the corresponding operators are the projectors onto a basis in this

[1] The term entanglement comes from the german term *Verschränkung*

space. More precisely, consider a basis in $\mathbb{C}^d$, a set of orthonormal vector $|i\rangle$ such that $\langle i|j\rangle = \delta_{ij}$[2]. A measurement in this basis is represented by the projectors $A_i = |i\rangle\langle i|$. Note that if the initial state is ρ, the probability of obtaining outcome i is:

$$p_i = \mathrm{tr}(|i\rangle\langle i|\rho|i\rangle\langle i|) = \langle i|\rho|i\rangle$$

while the state is mapped into

$$\rho_i = \frac{|i\rangle\langle i|\rho|i\rangle\langle i|}{tr(|i\rangle\langle i|\rho|i\rangle\langle i|)} = |i\rangle\langle i|$$

For the pure state case, $\rho = |\psi\rangle\langle\psi|$, these expression become:

$$p_i = tr(|i\rangle\langle i||\psi\rangle\langle\psi||i\rangle\langle i|) =| \langle i|\psi\rangle \,|^2$$

and

$$\rho_i = \frac{|i\rangle\langle i||\psi\rangle\langle\psi||i\rangle\langle i|}{| \langle i|\psi\rangle \,|^2} = |i\rangle\langle i|$$

and thus,

$$|\psi_i\rangle = |i\rangle$$

that is, the initial state $|\psi\rangle$ collapses into $|i\rangle$ with probability given by the square of the overlap $|\langle i|\psi\rangle|^2$. The previous formula also implies that there is no measurement (physical process) distinguishing $|\psi\rangle$ from $e^{i\gamma}|\psi\rangle$. This is: the state of a physical system is actually described by a vector in a Hilbert space up to an irrelevant global phase.

2.2 The qubit

Once the tools needed for the analysis of any Quantum Information Theory application are defined, the first step is to introduce the basic unit of Quantum Information, the quantum bit or qubit. Consider the encoding of a classical bit into a quantum particle. The way of achieving is employing a two-dimensional system: bit value 0 will be encoded into a state $|\psi_0\rangle$, or simply $|0\rangle$ and 1 into $|1\rangle$. Since these two options have to define a classical bit, the states have to be orthogonal, so they define a basis in $\mathbb{C}^2$. Therefore, a measurement in this basis distinguishes in a deterministic way between the two possibilities, as it happens for a classical bit. For instance

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

A two-dimensional particle can encode a classical bit and then defines a quantum bit. Note however that because of the superposition principle, any coherent combination of $|0\rangle$ and $|1\rangle$ is also allowed, something that is impossible in classical terms. Therefore:

Definition 2.1 *A **qubit** is an element of $\mathcal{H} = \mathbb{C}^2$ of the form $a|0\rangle + b|1\rangle$, where $a, b \in \mathbb{C}$ such that $|a|^2 + |b|^2 = 1$*

It is possible to represent the state of a qubit by means of the so-called Poincaré or Bloch Sphere (see Figure 2.1). Using the fact that the global phase of any pure state is irrelevant , α can always be real. Then, any quantum bit can be specified by a complex number β, as α is fixed due to normalization. Thus, any state has the form:

$$|\psi\rangle = \begin{pmatrix} \cos(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2})e^{i\phi} \end{pmatrix}$$

[2] δ_{ij} denotes Kronecker's delta that is defined as $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise.

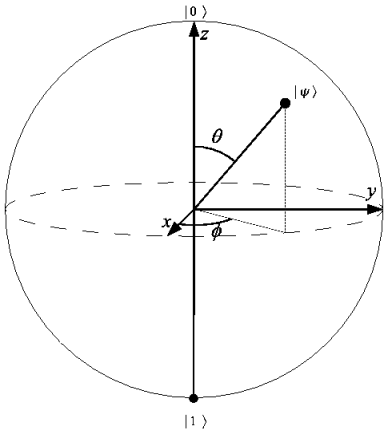


Figure 2.1: A quantum bit can be represented by a point in the surface of the Bloch Sphere. the corresponding unit vector $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ is called Bloch vector. Mixed states are inside the sphere.

Two angles θ and ϕ can specify any pure state in $\mathbb{C}^2$ as any point in the surface of the unit sphere. It is important to note that the orthogonal vector $|\psi_\perp\rangle$ is expressed as

$$|\psi_\perp\rangle = \begin{pmatrix} \sin(\frac{\theta}{2}) \\ -\cos(\frac{\theta}{2})e^{i\phi} \end{pmatrix}$$

Therefore, orthogonal vectors in $\mathbb{C}^2$ are represented by antiparallel vectors on the Bloch sphere.

The Pauli matrices play a significant role in the algebra of qubits and are important to introduce the quantum information processing applications that will follow.

The expression of Pauli matrices is:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The basis $|0\rangle, |1\rangle$ that define a qubit is formed by the eigenvectors of σ_z with eigenvalues +1 and -1 respectively, since $\sigma_z|n\rangle = (-1)^n|n\rangle$ for $n = 0, 1$. The effect of the other Pauli matrices in the qubit basis is:

$$\sigma_x|0\rangle = |1\rangle \quad \sigma_x|0\rangle = |1\rangle$$
$$\sigma_y|0\rangle = i|1\rangle \quad \sigma_y|1\rangle = -i|0\rangle$$

The effect of σ_x is a bit flip. The eigenvectors of σ_x and σ_y are obtained substituting respectively $(\theta, \phi) = (\pi/2, 0)$ and $(\theta, \phi) = (\pi/2, \pi/2)$ in the Bloch representations of $|\psi\rangle$ and $|\psi_\perp\rangle$. The Pauli matrices can also be expressed as

$$\sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0| \quad \sigma_y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| \quad \sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$$

Suppose a generic pure state of a qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. The associated projector is therefore

$$|\psi\rangle\langle\psi| = \mid \alpha \mid^2 |0\rangle\langle 0| + \mid \beta \mid^2 |0\rangle\langle 0| + \alpha\beta^*|0\rangle\langle 1| + \alpha^*\beta|1\rangle\langle 0|$$

Recalling the expressions of Pauli matrices in terms of projectors, one has:

$$|\psi\rangle\langle\psi| = \frac{1}{2}(\mathbb{1} + (\mid \alpha \mid^2 - \mid \beta \mid^2)\sigma_z + 2\,\mathrm{Re}(\alpha\beta^*)\sigma_x + 2\,\mathrm{Im}(\alpha\beta^*)\sigma_y)$$

From the fact that $\sigma_k^2 = \mathbb{I}$ it follows that $|\psi\rangle\langle\psi| = \frac{1}{2}(\mathbb{1} + \sum_k \langle\sigma_k\rangle_\psi \sigma_k)$, that can be expressed as:

$$|\psi\rangle\langle\psi| = \frac{1}{2}(\mathbb{1} + \hat{n}\cdot\vec{\sigma}), \text{ with } \hat{n} = \begin{pmatrix} \langle\sigma_x\rangle_\psi \\ \langle\sigma_y\rangle_\psi \\ \langle\sigma_z\rangle_\psi \end{pmatrix} = \begin{pmatrix} 2\,\mathrm{Re}(\alpha\beta^*) \\ 2\,\mathrm{Im}(\alpha\beta^*) \\ \mid \alpha \mid^2 - \mid \beta \mid^2 \end{pmatrix}$$

The vector $\hat{n}$ corresponds to the expectation value of the "magnetic moment" $\vec{\sigma}$ in the given state. It is expressed in terms of the Bloch sphere, already discussed. Indeed there is a one-to-one correspondence employing spherical coordinates with the expressions obtained in the previous section:

$$|\psi\rangle \equiv |+\hat{n}\rangle = \cos\frac{\theta}{2}|0\rangle + \exp i\phi \sin\frac{\theta}{2}|1\rangle \longleftrightarrow \hat{n} \equiv \hat{n}(\theta,\phi) = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}$$

A two-dimensional particle, such as a spin-$\frac{1}{2}$ particle can encode a classical bit: it is enough to prepare two states of a given basis and always measure in this basis. However, it is possible to do more than this as prepare coherent superpositions and measure in different basis. A quantum bit is potentially richer than a classical bit (cbit). The first question one can raise is how many cbits are required to transmit a qubit. A generic qubit is determined by a complex number and would need infinite bits to be encoded. But this doesn't mean that one qubit contains an infinite amount of classical information: the information encoded in a quantum state has to be read through a measurement. Only one cbit is accessible to the observer once the measurement is made. This fact is true for any amount of qubits. Although the amount of information accessible in an encoded quantum particle is the same as in a classical encoding, employing quantum states one can accomplish tasks that are impossible in terms of classical information theory.

2.3 Entanglement and Bell inequalities

Before giving some examples of applications of quantum information processing, let's describe more deeply the feature of entanglement which is sometimes described as the most quantum feature. It is a physical property with relevance in Quantum Information Theory and has a very important role in understanding Quantum Mechanics. This is because all entangled pure states violate the so-called Bell inequalities which are experimentally measurable conditions that allow to test quantum mechanics against the set of local realistic models introduced by Einstein, Podolsky and Rosen (EPR) in 1935.

In 1935, these three authors published a letter entitled "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?", where they raised doubts about the completeness of QM as local-realistic theory. Their arguments were for a state in an infinite-dimensional composite system, but can easily be adapted to the simplest case of two qubits.

First of all, they presented three requirements any physical theory should satisfy:

- Locality: Events in space-like separated regions cannot have causal relation.
- Reality: If, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.
- Completeness: Every element of the physical reality must have a counterpart in the physical theory.

All these conditions seem intuitively natural and hardly restrictive. Let's apply these requirements to a quantum scenario. Consider two separate parties sharing an entangled singlet state,

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

When both parties, Alice and Bob, measure in the computational basis, or z direction, they obtain perfectly anti-correlated outcomes. Actually, this also happens for all possible bases. Denoting by $|\pm\hat{n}\rangle$ any vector in the Bloch sphere and its orthogonal, one can see that

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|+\hat{n}\rangle|-\hat{n}\rangle - |-\hat{n}\rangle|+\hat{n}\rangle)$$

which follows from the fact that the singlet is rotationally invariant, it does not change when the same unitary transformation is applied on both sites, i.e. $|\Psi^-\rangle = U \otimes U|\Psi^-\rangle$. This implies that perfect anti-correlation are observed for all bases. Therefore, when Alice measures in the direction $\hat{n}$ and obtains the result $+(-)$, she knows the outcome of Bob's measurement in the same direction will be $-(+)$. Imagine now an experiment where a source is preparing and sending a singlet state to two distant parties. The parties choose between two non-commuting observables, say σ_x and σ_z. The choices by Alice and Bob are such that they define space-separated event, i.e. they cannot influence each other. If QM has to be consistent with the three previous postulates, it is necessary that:

- After reading the outcome for her measurement, say z, Alice knows in a deterministic way what the outcome of the same measurement by Bob is.

- Because of the space-like separation of events between the parties and relativity, Alice's choice of measurement cannot influence Bob's result. Alice is able to predict the value of σ_z for Bob without disturbing his system. Therefore σ_z on Bob's side is an element of physical reality. But one could have applied the same reasoning in the case Alice chooses σ_x and again conclude that σ_x for Bob is another element of physical reality.

- These two elements of physical reality should be reflected by the physical theory. However, QM cannot assign definite values to σ_x and σ_z because they are non-commuting observables. Then, QM is not a complete theory.

The conclusion of the EPR program was that there should be a complete theory alternative to QM.
J.Bell in 1964 ([10]) identified a series of conditions, the so-called Bell inequalities, that any local realistic theory should satisfy and provided a quantum experiment that violated the inequalities. The experimental demonstration of Bell inequalities violation closed the EPR program for the existence of a local realistic theory alternative to QM. In the following lines, we present the most known Bell inequality, the Clauser-Horne-Shimoni-Holt (CHSH) inequality, that will play a crucial role in the protocol presented in chapter 5.

The CHSH inequality refers to a scenario where two separated parties can choose between two measurements of two outcomes on two particles they receive from a common source (see Fig. 2.2). In a local theory, the outcome of the measurement by one of the parties, say Alice, has to be independent of Bob's measurement. It can only depend on the choice by Alice, x, and the common preparation at the source, denoted by λ. Therefore, if one denotes by a_x (b_y) the outcome for measurement A_x (B_y), one has $p(a_x|A_x, B_y, \lambda) = p(a_x|A_x, \lambda)$. In this simplest scenario, any deterministic preparation at the source should specify the outcomes for the four possible measurements. An example of such a preparation can be $\lambda = (+, +; -, +)$, i.e. $a_0 = a_1 = b_1 = 1$ and $b_0 = 0$. It is simple to see that for all these preparations

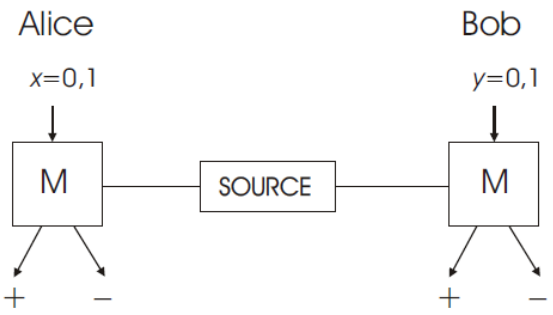


Figure 2.2: Experimental depiction of the CHSH inequality. Two separated parties measure 2 2-outcome observables, A_0 and A_1 for Alice and B_0 and B_1 form Bob. The choice of measurement is represented by a bit, x and y for Alice and Bob respectively. The two possible outcomes are labeled ± 1

$$B = a_0(b_0 + b_1) + a_1(b_0 - b_1) = \pm 2$$

Indeed since the outcomes are labeled by $\pm$, one has two possibilities:

1. $b_0 = b_1$ and then $B = 2a_0b_0 = \pm 2$ or
2. $b_0 = -b_1$ and then $B = 2a_1b_0 = \pm 2$ again.

In the more more general situation where the source is sending different preparations, λ, with some probability $p(\lambda)$, one can bound the expectation value of the previous quantity

$$-2 \leq \langle B \rangle = \langle a_0(b_0 + b_1) + a_1(b_0 - b_1) \rangle \leq 2$$

where $\langle\rangle$ denotes expectation with probability $p(\lambda)$. This is a condition that all theories with the restrictions established by EPR should satisfy.

It is relatively easy to design a quantum experiment where this inequality is violated. Consider the situation where the source is sending a pair of spin-one-half particles in the singlet state. Employing the algebra of Pauli matrices defined above, we have the following result: if Alice and Bob measure in the $\hat{n}_A$ and $\hat{n}_B$ directions respectively, then $\langle \sigma_{\hat{n}_A} \otimes \sigma_{\hat{n}_B} \rangle_{\Psi^-} = -\hat{n}_A \cdot \hat{n}_B$.

Doing some calculations, it is easy to see that

$$\vec{\sigma} \otimes \mathbb{1} |\Psi^-\rangle = -\mathbb{1} \otimes \vec{\sigma} |\Psi^-\rangle$$

Therefore,

$$\begin{aligned} \langle \sigma_{\hat{n}_A} \otimes \sigma_{\hat{n}_B} \rangle_{\Psi^-} &= \langle \Psi^- | (\sigma_{\hat{n}_A} \otimes \sigma_{\hat{n}_B}) | \Psi^- \rangle = -\langle \Psi^- | (\hat{n}_A \cdot \vec{\sigma})(\hat{n}_B \cdot \vec{\sigma})) \otimes \mathbb{1}) | \Psi^- \rangle = \\ &= -n_{Ai} n_{Bj} \operatorname{tr}(\rho_A \sigma_i \sigma_j) = -\hat{n}_A \cdot \hat{n}_B \end{aligned}$$

And that the density operator of Alice system can be expressed as $\rho_A = \frac{1}{2}\mathbb{1}$

The projector of the singlet state can be expressed as

$$P_{\Psi^-} = \rho = |\Psi^-\rangle\langle\Psi^-| = \frac{1}{4}(\mathbb{1} \otimes \mathbb{1} - \sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y - \sigma_z \otimes \sigma_z)$$

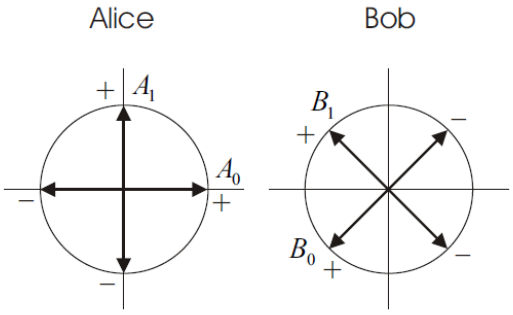


Figure 2.3: Measurement setting for Alice and Bob in the CHSH scenario

If Alice and Bob apply the measurements in Figure 2.3 considering that $\vec{a}_i = \hat{n}_{A_i} \cdot \vec{\sigma}$ and the previous expression $\langle \vec{a}_i \vec{b}_j \rangle = -\hat{n}_{A_i} \cdot \hat{n}_{B_j} = -\cos(\theta)$ where θ is the angle between the vectors,

$$\langle \vec{a}_0 \vec{b}_0 \rangle = \langle \vec{a}_0 \vec{b}_1 \rangle = \langle \vec{a}_1 \vec{b}_0 \rangle = -\cos(\frac{3\pi}{4}) = \frac{1}{\sqrt{2}}$$

and

$$\langle \vec{a}_1 \vec{b}_1 \rangle = -\cos(\frac{\pi}{4}) = -\frac{1}{\sqrt{2}}$$

So we have that the value of the expected value $\langle B \rangle_{\Psi^-}$ given a singlet is distributed and measured according to Figure 2.3 is:

$$\langle B \rangle_{\Psi^-} = \langle \vec{a}_0 \vec{b}_0 \rangle + \langle \vec{a}_0 \vec{b}_1 \rangle + \langle \vec{a}_1 \vec{b}_0 \rangle - \langle \vec{a}_1 \vec{b}_1 \rangle = 4\frac{1}{\sqrt{2}} = 2\sqrt{2}$$

This value is greater than 2, which means that Quantum Mechanics violates the condition that all EPR theories satisfy. There is no Local Realistic theory able to reproduce all quantum predictions, so either QM is wrong or the EPR program is not possible. Bell was then able to map all the EPR debate into a measurable condition. The only thing left was to design the experimental situation for testing the violation of any of these quantities. The rigorous experimental verification of the violation of a Bell inequality came almost two decades after Bell seminal paper. Since then, other experimental tests have been performed, all in favor of the quantum formalism.

Two facts should be mentioned that are important for the rest of development of the present document:

- First, although only one inequality is here presented, there are many similar conditions that characterize the set of probability distributions achievable in a local realistic theory, that read

  $$P(a_x, b_y \mid A_x, B_y, \lambda) = \sum_{\lambda} P(\lambda) P(a_x \mid A_x, \lambda) P(b_x \mid B_x, \lambda)$$

  The fact that quantum mechanics predicts a violation of a Bell inequality means that one cannot write a vector of quantum probabilities violating an inequality in the previous form (as a convex sum of separable correlations).

- Second, despite the fact that QM predicts the violation of these inequalities, no faster-than-light communication is allowed in the quantum formalism. This would be the case if Alice's measurement could change Bob's measurement statistics.

  It is straightforward to prove that this is not possible. Consider the situation where Alice and Bob share a quantum state ρ_{AB} on which they apply two measurements. These measurements are described by a set of positive operators, $\{A_x\}$ and $\{B_y\}$, that sum up to the identity. The partial state ρ_A is defined below, for any measurement A_x on Alice's particle. It must hold that

  $$\begin{aligned} \mathrm{tr}_A(\rho_A A_x) &= \mathrm{tr}_{AB}(\rho_{AB}(A_x \otimes \mathbb{1}_B)) = \\ &= \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \langle a_j, b_k | \rho_{AB}(A_x \otimes \mathbb{1}_B) | a_j, b_k \rangle = \\ &= \sum_{j=1}^{d_A} \langle a_j | \,\mathrm{tr}_B(\rho_{AB}) A_x | a_j \rangle = \mathrm{tr}_A(\mathrm{tr}_B(\rho_{AB}) A_x) \end{aligned} \tag{2.2}$$

  So by identification

  $$\rho_A = \mathrm{tr}_B(\rho_{AB}) = \sum_{k=1}^{d_B} \langle b_k | \rho_{AB} | b_k \rangle \tag{2.3}$$

  This is the general definition of the partial state on A which is obtained by partial trace over the other system B. Since the trace is an unitary invariant, it implies that whatever Bob does, the partial state of Alice will remain unchanged.

  The probability for Alice obtaining outcome x is

  $$P(x) = \sum_y P(x, y) = \sum_y \mathrm{tr}(A_x \otimes B_y \rho_{AB}) = \mathrm{tr}(A_x \otimes \mathbb{1}_B \rho_{AB}) = \mathrm{tr}(A_x \rho_A)$$

where we used that $\sum_y B_y = \mathbb{1}_B$ This means that Alice's local measurement statistics cannot be affected by any measurement by Bob (and viceversa), so no faster-than-light communication is possible. This is known as the no-signaling through entanglement principle, Bob cannot use entanglement to send any message to Alice.

As already stated, the concept of entanglement arises from the structure of quantum systems in Hilbert spaces. Considering a multipartite system consisting of n subsystems, the total state space is described in classical terms as the cartesian product of the n subsystems. This implies that the total state is always a product state of the n separate systems. Notwithstanding in quantum terms, the total Hilbert space H is a tensor product of the n subsystems, $H = \otimes_{l=1}^{n} H_l$. Then, the superposition principle allows to write the total state of the system as:

$$|\psi\rangle = \sum_{i_1,i_2\ldots,i_n} c_{i_1} \cdot c_{i_2} \cdots c_{i_1} |i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle = \sum_{\mathbf{i}_n} c_{\mathbf{i}_n} |\mathbf{i}_n\rangle \tag{2.4}$$

The total state of the system cannot be, in general, described as a product of states of individual subsystems, i.e. $|\psi\rangle \neq |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$. This means, that it is in general not possible, to assign a single state vector to any of n subsystems. It expresses formally a phenomenon of entanglement, which in contrast to classical superposition, allows to construct exponentially large superposition with only linear amount of physical resources. It is just what allows to perform nonclassical tasks employing quantum features of nature.

In the case of a bipartite pure system, all the entanglement properties of a given state $|\psi\rangle \in C^{d_A} \otimes C^{d_B}$ can be derived from its Schmidt decomposition, obtained from the singular value decomposition of the matrix of coefficients of the bipartite state in a given basis. Given the bases $\{|j\rangle_A\}$ and $\{|k\rangle_B\}$, then the state can be written as:

$$|\psi\rangle = \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} c_{jk} |j\rangle_A |k\rangle_B \tag{2.5}$$

And the matrix C of the coefficients c_{ij} can be decomposed in the form $C = UDV$ where D is diagonal, with $D_{ij} = \lambda_i \delta_{ij}$ and U and V are unitary matrices. The bipartite state can be expressed in the orthonormal Schmidt bases $|i\rangle_A$ and $|i\rangle_B$.

$$|\psi\rangle = \sum_i \lambda_i |i\rangle_A |i\rangle_B = \sum_i \lambda_i |ii\rangle \tag{2.6}$$

The set of $\{\lambda_i\}$ are called Schmidt coefficients. The Schmidt decomposition allows us detecting when a given pure state is entangled: That is when its Schmidt decomposition has more than one term.

Entanglement of mixed states is not longer equivalent to being non-product, as in the case of pure states. Instead, one calls a mixed state of n systems entangled if it cannot be written as a convex combination of product states. In classical terms, a probability distributions can always be written as mixtures of product distributions.

$$\rho \neq \sum_i p_i \rho_1^i \otimes \cdots \otimes \rho_n^i \tag{2.7}$$

The states which are not entangled are called separable. In practice it is hard to decide if a given states is separable or entangled basing on the definition itself. Therefore one of the fundamental problems concerning entanglement is the so called separability problem. We can define entangled states as the ones that cannot be simulated by classical correlations.

This section, apart from its fundamental and historical importance, shows why entanglement is so important in Quantum Mechanics and Quantum Information Theory. It is a concept that cannot be explained by a standard (local) model. Quantum correlations have no classical analog, so one can hope to find new information tasks exploiting these intrinsically quantum features. In the following sections, we show some of these applications.

2.3.1 A note on experiments and loopholes

Two possible loopholes have been identified in the experimental scrutinizing of quantum non-local correlations.

- If one does not arrange the timing adequately, the detections may be attributed to sub-luminal communication: this is called the *locality loophole*. In order to close this loophole, the events must be ordered in space time in such a way that arrival and measurement of each particle on Alice's and Bob's side must be performed before the light cone generated by the moment of measurement choice arrives the other party.
- The other possible problem is known as *detection loophole*. In all experiments, the violation of Bell's inequalities is measured on the events in which both particles have been detected. Since detectors don't have perfect efficiency, there is also a large number of events in which only one particle is detected, while the other has been missed. The detection loophole assumes a form of conspiracy, in which the undetected particle "chose" not to be detected after learning to which measurement it was being submitted. In this scenario, if the detection efficiency is not too high, it is pretty simple to produce an apparent violation Bell's inequalities with local variables.

2.3.2 Local operations with classical communications

The set of local operations assisted with classical communication (LOCC) plays a fundamental role in the characterization and quantification of entangled states. The only way these correlations can be created between two separated parties is

1. by sending quantum particles though a quantum channel or
2. meeting together and performing a global quantum operation.

If these two operations are not possible, it is clear that no entanglement can be created. In other words, if two parties, Alice and Bob, share a quantum state, the amount of entanglement in this state cannot be increased if they exchange classical communication or perform operations on their local systems. It is then said that Entanglement is a quantum resource that cannot increase under LOCC.

Assume now one has to compare two states, ψ_1 and ψ_2, in terms of their entanglement properties. If ψ_1 can be transformed into ψ_2 by LOCC, then ψ_1 is at least as entangled as ψ_2. Indeed, any Quantum Information processing task by Alice and Bob using ψ_2 can be performed if they start from ψ_1. Consequently, one can say that the study of the entanglement properties of quantum states, then, can be rephrased as the study of their interconversion by LOCC. The following section gives examples of these tasks.

2.4 Some examples of Quantum information theory

This section exposes different applications of quantum mechanics to information processing tasks, concentrating on communication applications where separated parties send information encoded on quantum states.

First, dense coding and quantum teleportation are introduced as examples of intrinsically quantum information tasks without classical analog. Then the quantum no-cloning theorem is derived and is used to introduce Quantum Cryptography.

2.4.1 Dense coding

Suppose that two separate parties, Alice and Bob, are willing to interchange information through quantum states. It is known that if Alice can only send a qubit, she cannot transmit more than one classical bit. If Alice and Bob share the following entangled state:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0_A\rangle|0_B\rangle + |1_A\rangle|1_B\rangle)$$

Alice and Bob's quantum systems are correlated, they share a maximally entangled state of two qubits (called 1 entangled bit). Alice wants to send classical information to Bob by sending a quantum particle. As it is shown below, she can indeed transmit 2 classical bit by sending her half of the shared state. Alice applies on her local state one of the following unitary operations: $\mathbb{1}$, σ_x, $i\sigma_y$ or σ_z. The resulting states are:

$$\begin{aligned}
|\Phi^+\rangle &= (\mathbb{1}\otimes\mathbb{1})|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
|\Psi^-\rangle &= (i\sigma_y\otimes\mathbb{1})|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \\
|\Psi^+\rangle &= (\sigma_x\otimes\mathbb{1})|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\
|\Phi^-\rangle &= (\sigma_z\otimes\mathbb{1})|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)
\end{aligned} \tag{2.8}$$

As already discussed in 2.1, these four states form a basis in $\mathbb{C}^2\otimes\mathbb{C}^2$, the so-called Bell basis (Eq. 2.1). Now, after its processing, Alice sends her particle to Bob. Bob has to distinguish between four orthogonoal two-qubit states, so he only has to measure in the Bell basis. He can then read the two bits of information encoded by Alice in a deterministic way. These steps are the so-called dense coding protocol.

The sharing of an entanglement state at the beginning of the protocol is what allows to communicate 2 bits sending one qubit. This protocol is not useful in a practical way, but its fundamental interest is in showing how entanglement is at the core of a lot of quantum information processing tasks. This can be expressed in the following terms[3]:

$$1\,ebit + 1\,qubit \rightarrow 2\,cbits$$

2.4.2 Quantum teleportation

Quantum teleportation can be seen as the process complementary to dense coding but is a much more important communication scheme, from a fundamental and applied point of view.

The scheme assumes as usual two separate parties. Alice receives an unknown qubit, e.g. a spin-one-half particle pointing into an unknown direction. She has to transmit this particle to Bob. Assume she cannot transmit the particle directly because the channel is noisy and quantum information can't be reliably transmitted through it. Alice could try to read her quantum state and send the information acquired by the measurement to Bob, who could prepare a quantum state accordingly. However, this procedure is invalid: Alice's state is unknown and her measurement will perturb it. Additionally, if her measurement could provide her full information about her state, she would require an infinite number of bits to describe it to Bob and help him prepare the state (in general). If Alice and Bob share at the beginning a maximally entangled state of two qubits, as in the dense coding situation, they can improve their situation. Denoting by P the particle in the unknown state $|\psi\rangle$ to be transmitted, where $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the global initial state reads:

$$|\Psi\rangle = |\psi\rangle_P|\Phi^+\rangle_{AB}$$

After some algebra, one can rewrite the previous state as:

$$\begin{aligned}
|\Psi\rangle =& \frac{1}{2}(|\Phi^+\rangle_{PA}(\alpha|0\rangle + \beta|1\rangle)_B + |\Phi^-\rangle_{PA}(\alpha|0\rangle - \beta|1\rangle)_B + \\
&+ |\Psi^+\rangle_{PA}(\alpha|1\rangle + \beta|0\rangle)_B + |\Psi^-\rangle_{PA}(\alpha|1\rangle - \beta|0\rangle)_B)
\end{aligned} \tag{2.9}$$

[3] *ebit* stands for pair of entangled bits and *cbit* stands for classical bit

Alice then measures employing the Bell basis defined in equation 2.1. Bob's state changes depending on Alice's result. For instance, if Alice projects her two-qubit state onto $|\Psi^+\rangle$, Bob's state is $\alpha|1\rangle + \beta|0\rangle$. For all possible results from Alice's measurement, Bob's state resembles the initial state of P. Alice communicates the result to Bob and since the four possible outcomes have the same probability, she has to send two bits to identify her result. Depending on the information he receives, Bob then applies a unitary operation that transforms his state into $|\psi\rangle$ as:

$$\begin{aligned}\mathbb{1}((\alpha|0\rangle + \beta|1\rangle)_B) &= |\psi\rangle \\ \sigma_z((\alpha|0\rangle - \beta|1\rangle)_B) &= |\psi\rangle \\ \sigma_x((\alpha|1\rangle + \beta|0\rangle)_B) &= |\psi\rangle \\ \sigma_z\sigma_x((\alpha|1\rangle - \beta|0\rangle)_B) &= |\psi\rangle\end{aligned}$$

After applying the corresponding operation, Bob's particle is in P's unknown state.

It is important to notice that no quantum particle is sent through the channel. Indeed, the two bits Alice sends do not contain information on $|\psi\rangle$. The quantum information on the state travels through the correlations on $|\Phi^+\rangle$. This does not mean that two classical bits are sufficient to send all the information contained on a quantum state; the process of quantum teleportation is impossible if the parties don't share an entangled state. After the information processing, no correlations are left between Alice and Bob (their final state is product). It is said that entanglement is consumed. The process here presented does not contradict Einstein's special relativity, there is no faster than light communication. Bob's must receive Alice's information on her measurement result in order to correct his quantum state choosing which unitary transformation apply. Before acquiring this information, Bob's state is a mixture of the four states described, which gives a complete noisy state ($\rho_B = \frac{1}{2}\mathbb{1}$). Finally, once the task is completed, Alice does not keep any information on the initial state $|\psi\rangle$. So, teleportation is consistent with the following property of Quantum Information, the no-Cloning theorem. It can be expressed as:

$$1\,ebit + 2\,cbits \rightarrow 1\,qubit$$

2.4.3 Quantum No Cloning Theorem

The previous sections have shown Quantum Information applications that exploit quantum effects with no classical analog. Quantum teleportation also improves the understanding of the coexistence of Quantum Mechanics and Special Relativity. However, the encoding of information on quantum states suffer from limitations that do not appear in classical information theory. The most important one, here presented, is given by the quantum no cloning theorem. It states that quantum information cannot be copied. At first sight this can be considered as a serious drawback for information processing. However there are ways to circumvent this problem (with quantum error correction techniques) and, moreover, this limitation can be turned in an advantage in the form of Quantum Cryptography. To prove the No-cloning theorem. let's first assume that there is a machine duplicating the quantum state of a system, i.e. for any incoming state $|\psi\rangle \in \mathbb{C}^2$, the machine outputs $|\psi\rangle|\psi\rangle$. In sake of simplicity, let's assume the system to be of dimension 2. This quantum process has to be mathematically described by a linear map $\mathcal{L}$

$$\mathcal{L}(|\psi\rangle \otimes |C\rangle) = |\psi\rangle \otimes |\psi\rangle$$

where $|C\rangle$ is the state of the machine where the clone is produced. Since it is assumed that the machine works for any initial state,

$$\begin{aligned}\mathcal{L}(|0\rangle \otimes |C\rangle) &= |0\rangle \otimes |0\rangle \\ \mathcal{L}(|1\rangle \otimes |C\rangle) &= |1\rangle \otimes |1\rangle\end{aligned} \tag{2.10}$$

Since $\mathcal{L}$ is linear,

$$\mathcal{L}(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |C\rangle) = \frac{1}{\sqrt{2}}(\mathcal{L}(|0\rangle \otimes |C\rangle) + \mathcal{L}(|1\rangle \otimes |C\rangle))$$
$$= \frac{1}{\sqrt{2}}(|0\rangle|0\rangle) + |1\rangle|1\rangle))$$

which is not equal to the searched state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$

Therefore, the linearity of Quantum Mechanics makes the cloning process impossible. One can construct a cloning machine producing two copies of two orthogonal states but it will fail when cloning any coherent superposition of these two states. Superposition principle and nonorthogonality are at the core of the no-cloning theorem.

2.5 Introduction to Quantum Cryptography

From the viewpoint of this memoir, all the previous sections are the necessary steps to arrive to this section, the description of the employment of quantum features to attain secrecy and achieve secure key distribution. The presentation and understanding of the quantum formalism, quantum measurement, entanglement and the quantum cloning theorem leads to the introduction of quantum cryptography as it was originally developed.

The previous section, shows limitations of quantum information processing but also opens the door to new possibilities. The No cloning theorem is connected to a more known limitation of the quantum formalism: the measurement process perturbs the state of the system. Indeed, perfect cloning would violate the fact that the state of a single quantum system cannot be perfectly known (producing a clones and measuring them, one could know everything about the original system without interacting with it). The perturbation of the state of a quantum system once one tries to measure it can be employed to detect the action of any spy trying to read quantum information propagating through a channel. Or in other words, the action of the spy, also known as eavesdropper, is limited by the impossibility of producing a perfect copy of the quantum state. These two ideas lie at the basis of any Quantum Cryptography protocol. Here, the first of these schemes, the so-called BB84 protocol, invented in 1984 by Bennett and Brassard ([7]), is presented.

2.5.1 BB84 protocol

In any cryptographic scenario, two honest parties, Alice and Bob, want to exchange information in a private way. There is also a third dishonest party, the eavesdropper, also called Eve, that wants to read this information. Alice and Bob's task in any standard cryptographic protocol for key distribution is to establish a shared secret key, namely a list of perfectly correlated bits which Eve has no information about. This secret key is later employed to encode information in a secure way. In the case of the BB84 protocol, the key is established as follows:

1. Alice chooses randomly one of the four possible states:

$$|\pm x\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) \quad |\pm y\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle) \tag{2.11}$$

   The bit 0 (1) is encoded onto $|+x\rangle$ and $|+y\rangle$ ($|-x\rangle$ and $|-y\rangle$). The vectors $|\pm x\rangle$ and $|\pm y\rangle$ define the x and y basis.

2. Bob measures randomly in one of the two bases. He maps his result into a classical bit using the same convention as Alice.

3. Alice broadcasts through a classical channel her basis preparation, x and y. In case Bob's measurement basis is the same as Alice's, he keeps the symbol, else he rejects. This process is called basis reconciliation.

Note that if Alice and Bob's bases agree, their bits are perfectly correlated. On the other hand, if the bases are different, Alice's preparation and Bob's result are completely uncorrelated. Indeed, assume that Alice has sent $|\pm x\rangle$ and Bob measures in the y basis. He will obtain the outcome corresponding to $|+y\rangle$ with probability $P(+y \mid +x) = |\langle +x|+y\rangle|^2 = \frac{1}{2}$. This is equivalent to Bob discarding Alice's state and flipping a coin. Therefore, after the basis reconciliation, Alice and Bob discard all the bad instances and share a list of perfectly correlated random bits.

Why Eve cannot break this protocol? In order to retrieve information, she has to interact with the quantum particle transmitted (say a photon), encoding the information as it is propagating to Bob By the No-Cloning theorem, she cannot make a perfect copy of the particle, forward the first clone to Bob and keep the second one.

However, she can try to measure the quantum state of the particle and read the information. At the moment she can make her measurement, she does not know the basis chosen by Alice, as it is transmitted later. This strategy is called intercept-resend. This simple and even practical attack consists in Eve measuring each qubit in one of the two bases, precisely as Bob does. Then, she resends to Bob another qubit in the state corresponding to her measurement result. In about half of the cases Eve will be lucky and choose the basis compatible with the state prepared by Alice. In these cases she resends to Bob a qubit in the correct state and Alice and Bob won't notice her intervention. However, in the other 50% cases, Eve unluckily uses the basis incompatible with the state prepared by Alice. This necessarily happens, since Eve has no information on Alice's random generator (hence the importance that this generator is truly random). Focusing on the cases where Bob's basis is the same as Alice's (after all, these are the only cases that pass the basis reconciliation process to generate a secret key), since Eve may have prepared a "wrong" state, Bob will obtain the expected result with probability one half. That is, Eve's strategy is introducing errors on Bob's side. Alice and Bob can detect Eve's intervention by making public a fraction of the symbols where their bases agree. If there are no errors, it is very likely that nobody has tried to eavesdrop their communication, and they can safely employed the remaining bits as a secret key. If they see an unexpectedly high amount of errors, someone is eavesdropping the channel, so they abort. Therefore, what Quantum Cryptography prevents is that an eavesdropper reads the information Alice and Bob are exchanging without being detected.
The next chapter introduces concepts from classical information theory that are able to increase secrecy in a quantum key distribution protocol.

2.5.2 EPR Protocol and entanglement as source of secrecy

Before finishing this brief introductory chapter about quantum information theory and cryptography, let's explore the relationship between entanglement and quantum cryptography.

The key bits generated in the BB84 protocol may appear to have been originated by Alice. However, it can be seen that the key can arise from a random process involving the properties of entanglement. The following protocol illustrates this and is the motivation of the protocol discussed ahead - that is a fundamental randomness with the same properties as entangled particles can be employed to extract a secret key-.
Alice could have prepared the states and send half of each to Bob, or a third party could have prepared the pairs of entangled states and send half to each of the parties. If we replace the quantum channel carrying qubits from Alice to Bob by a channel carrying 2 qubits from a common source, one qubit to Alice and one to Bob, a first possibility would be that the source emits the two qubits always in the same state chosen randomly among the 4 states of the BB84 protocol. Alice and Bob would then both measure their qubit in one of the two bases, again chosen independently and randomly. The source then announces the bases and Alice and Bob keep the data only when they happen to have done their measurements in the compatible basis. If the source is reliable, this protocol is equivalent to the BB84 one: Everything is as if the qubit propagates backwards in time from Alice to the source, and then forward to Bob. But better than trusting the source, which could be in Eve's hand, the EPR protocol assumes that the n qubits pairs are emitted in a maximally entangled state like:

$$|\Psi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

In his 1991 paper Artur Ekert suggested to base the security of this 2−qubit protocol on Bell's inequality. For this, Alice and Bob have a third choice of basis then select a random subset of their entangled pairs and test if they violate Bell's inequality (to establish that they are not product states). In this way the probability that they happen to choose the same basis is reduced from $\frac{1}{2}$ to $\frac{2}{9}$, but at the same time as they establish a key they collect enough data to test Bell inequality. If the test is passed, it means they hold sufficiently pure entangled states, placing a lower bound on the fidelity of the remaining pairs. Then, when Alice and Bob happen to use the same basis, both the $x-$basis or both the $y-$basis, i.e. in about half of the cases, their results are identical, providing them with a common key.

Although the exact relation between security and Bell inequality is not yet fully known, there are clear results establishing fascinating connections. The protocol analyzed in chapter 5 was a step forward to deepen the understanding of this relationship.

Chapter 3

Information theory for secret key agreement

This chapter provides the definitions and results from an information-theoretic security standpoint that will be employed in the no-signaling key distribution protocols of chapter 5. The concepts of advantage distillation, privacy amplification and intrinsic information are described.

One of the basic problems in cryptography is to transmit a message securely from a sender, Alice, to a receiver, Bob, over an insecure channel overheard by an adversary/eavesdropper, Eve, in such a way that Eve does not learn any information about the message. One possibility for solving this problem is that the parties generate a secret key by means of insecure communication and then use this key as the private key of a classical cryptosystem (e.g. a one-time pad) for transmitting in a secure way the message. This is the motivation behind the problem of secret key agreement by insecure communication.

Quantum cryptography has triggered since its origins research in security key agreement based in information theory. This will be the framework that we will employ to study the protocol presented in chapter 5. The security of an information-theoretically-secure scheme is based in statements from information theory, initiated in 1948 by Claude Shannon, the mathematical theory of information and communication based on probability theory and statistics.

As a sort of introduction to the rest of the chapter, let's give some details about the relationships between quantum cryptography and security information theory that are relevant for this project. The best known example of this relationship is probably the development of privacy amplification algorithms. This in turn triggered the development of new cryptosystems based on weak but classical signals, emitted for instance by satellites (Maurer 1993). These new developments required secret key agreement protocols that can be used even when Eve has more information than the honest parties. Such protocols, called advantage distillation, necessarily use two way communication and are much less efficient than privacy amplification. It is somewhat surprising that secret key agreement is possible even if Alice and Bob start with less mutual (Shannon) information than Eve. However, they can take advantage of the authenticated public channel: Alice and Bob can decide which series of realizations to keep, whereas Eve can't influence this process. (Maurer 1993, Maurer and Wolf 1999). A second remarkable connection between quantum and classical secret key agreement is: If Eve follows the strategy which optimizes her Shannon information, under the assumption that she attacks the qubit one at a time (the so-called individual attacks, that we will define in chapter 5), then Alice and Bob can use advantage distillation if and only if Alice and Bob's qubits are still entangled. This connection between the concept of entanglement, central to quantum information theory, and the concept of intrinsic classical information, central to classical information based cryptography (Maurer and Wolf 1999), has been shown to be general (Gisin and Wolf 2000). This chapter will deal with the definition of these concepts.

3.1 Classical Information theory and Entropy

For the rest of this chapter, we will use the following definitions and notation:

A discrete random experiment is a pair (Ω, P), where Ω is a finite or countably-infinite set, called

sample space and P is a function

$$P : \Omega \to [0, 1]$$

called the probability function, such that

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

An element $\omega \in \Omega$ is called an elementary event. A set $\mathcal{A} \subseteq \Omega$ is an event. The mapping P is extended to the set of parts of Ω, i.e. $\mathcal{P}(\Omega) = 2^{\Omega}$

$$P\left[\mathcal{A}\right] = \sum_{\omega \in (A)} P(\omega)$$

The properties of the probability function are the described, for example in [31].

X denotes a random variable with values in $\mathcal{X}$. Its associated probability distribution is defined as $P_X(x) = \sum_{\omega : X(\omega) = x} P(\omega)$.

Let's give definitions and properties about classical information theory.

Definition 3.1 *The (Shannon) entropy $H(X)$ of a random variable X with probability distribution P_X is*

$$H(X) \equiv -\sum_{x \in \mathcal{X}} P_X(x) \log(P_X(x)) = E\left[-\log P_X\right] \tag{3.1}$$

we denote by log *the binary logarithm.*

The units of the entropy are measured in bits. The entropy of a binary random variable with probability distribution $[p, 1-p]$ is given by the binary entropy function

$$h(p) \equiv -p \log p - (1-p) \log(1-p) \tag{3.2}$$

The joint entropy of random variables $X_1, X_2, \ldots, X_n$ is the entropy of the joint distribution $H(X_1 X_2 \cdots X_n) = H(P_{X_1 X_2 \cdots X_n})$

Proposition 3.1 *The entropy has the following properties*

$$0 \leq H(X) \leq \log |\mathcal{X}|$$

with equality in the following cases:

- $H(X) = 0$ *if and only if* $\exists x \in \mathcal{X}$ *tal que* $P_X(x) = 1$
- $H(X) = \log |\mathcal{X}|$ *iff* $\forall x \in \mathcal{X}$ *we have* $\mathrm{p}(x) = \frac{1}{|\mathcal{X}|}$, *i.e. X is uniformly distributed.*

For random variables X and Y, we have

$$H(XY) \leq H(X) + H(Y)$$

with equality if and only if X and Y are statistically independent. The quantity $H(XY) - H(X)$ can be interpreted as the entropy of the random variable Y when X is given.

Definition 3.2 *The conditional entropy of Y when given X is defined as*

$$H(Y|X) \equiv H(XY) - H(X) \tag{3.3}$$

This entropy is not the entropy of a specific probability distribution, but the expectation of entropies $H(Y|X = x)$

$$H(Y|X) \equiv E_X\left[H(Y|X = x)\right] \tag{3.4}$$

The conditional entropy satisfies

$$0 \leq H(Y|X) \leq H(Y) \tag{3.5}$$

This inequality can be intepreted as the fact that additional knowledge can never increase uncertainty. We define the quantity

Definition 3.3 *The mutual information between X and Y is defined as:*

$$I(Y;X) \equiv H(Y) - H(Y|X) = H(X) + H(Y) - H(XY) \geq 0 \tag{3.6}$$

The following holds

$$I(Y;X) = I(X;Y)$$

Analogously, the quantity $I(Y;X|\mathcal{A}) \equiv H(X|\mathcal{A}) - H(X|Y,\mathcal{A})$ can be defined

Definition 3.4 *The conditional mutual information between X and Y, given Z is defined as:*

$$I(X;Y|Z) \equiv H(X|Z) - H(X|Y,Z) = E_Z\left[I(X;Y|Z=z)\right] \tag{3.7}$$

The Figure 3.1 represents the different quantities defined for two random variables in a graphical way.

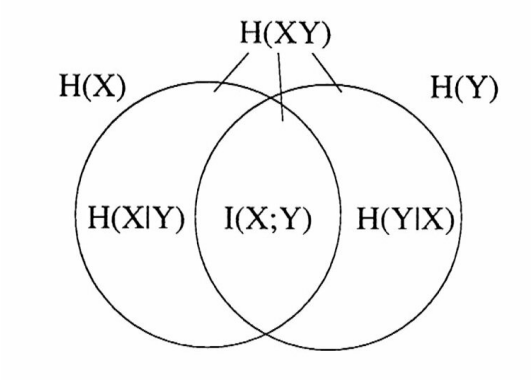


Figure 3.1: Graphical representation of mutual information for two random variables

For three random variables, the graphical representation is shown in figure 3.2.

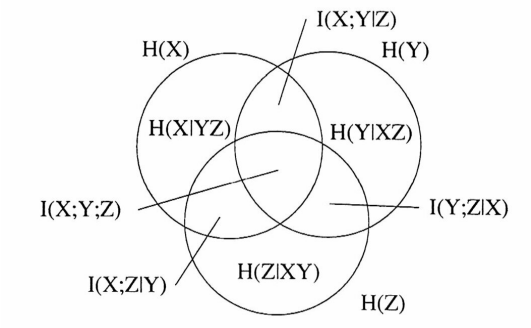


Figure 3.2: Graphical representation of mutual information for three random variables and respective conditional mutual information.

The quantity represented in the region in the middle

$$I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$$

is symmetric in X, Y and Z and can be negative. the other regions represent information-theoretic quantities that are always positive.

This brief description of the definitions from information theory is a prerequisite to the results of unconditionally secure key agreement of the following sections.

3.2 Unconditionally secure key agreement

The field of information theoretic perfect secrecy and unconditionally secure key agreement began with the publication of Shannon's seminal paper [20]. The result of perfect secrecy and the impracticality of secret key extraction derived from Shannon's theorem was subsequently averted due to modifications in the models and security requirements to achieve practical information theoretic security in realistic scenarios.

The classical scenario of symmetric cryptosystems is as follows: a message M, a key K and a ciphertext C; and the trusted parties Alice and Bob, and the attacker Eve.

Definition 3.5 *A cipher is called perfectly secret if the ciphertext reveals no information about the message, that is, if* $I(M;C) = 0$.

The one-time pad which consists in the use of a uniformly distributed key independent of M and with its same length is perfectly secure. The ciphertext is obtained as $C = M \oplus K$. To prove this, note that the key is uniquely determined when given the message an ciphertext, i.e. $H(K, MC) = 0$. Additionally, we have $I(K;C|M) = H(K|M) - H(K|CM) = H(K) + I(K;M) = N$, being N the message length. It follows that $I(M;C) = H(C) - H(C|M)$ and since $H(C) \leq \log|\mathcal{C}| = N$ and $H(C|M) = H(C|KM) + I(C;K|M)$ we have $0 \leq I(M;C) \leq 0$.

For perfect secrecy, a secret key as long as the message must be shared and used once. Shannon's theorem asserts the optimality of this cryptosystem respect to key length.

Theorem 3.1 *For every perfectly secret uniquely decodable cryptosystem we have*

$$H(K) \geq H(M)$$

This theorem implies that perfect secrecy is possible only between parties who share a secret key of length at least equal to the entropy of the message transmitted. Thus every perfectly secret cipher has the impracticality of the one time pad. However, the assumption that the adversary has perfect access to the ciphertext is too pessimistic and unrealistic. Thus analyzing secret agreement models in which the information the adversary obtains is limited in some way, translates into deriving information-theoretically-secure key agreement and secret message transmission. The condition of the existence of a bound in the adversary's knowledge can be achieved on the presence of noise in the communication channel, the fact that the adversary's memory capacity is limited or on the principles of quantum mechanics.

The general scenario for an interactive model of secret key agreement by public discussion and common information is described as:

- The parties Alice and Bob who want to establish a mutual secret key have access to the realizations of random variables A and B respectively.
- The adversary Eve knows a random variable E
- P_{ABE} is the joint distribution of the random variables.
- The legitimate partners are connected by an unsecure but authentic channel: it can be passively overheard by Eve but over which no undetected active attacks are possible.

Assuming that the parties have access to a number of independent realizations of the corresponding random variables, the secret key rate in this model is defined as the maximal rate at which Alice and Bob can generate a highly secret key by communication over the insecure channel, depending on the amount of randomness (number of realizations of X and Y) necessary. The secret key rate is denoted by $S(A;B \parallel E)$.

Let's describe each of the phases of secret-key-agreements protocols with a realistic scenario described in figure 3.3. A source sends out random bits and the three parties Alice, Bob and Eve receive this bits over independent binary symmetric channels with error probabilities ϵ_A, ϵ_B and ϵ_E respectively[1].

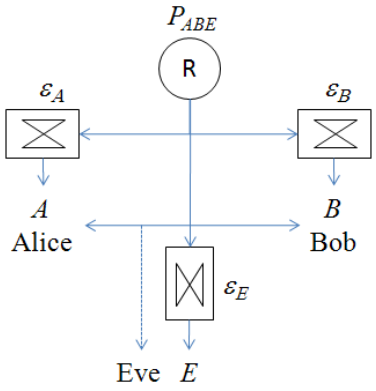


Figure 3.3: Secret key agreement by public communication from common information: the source of randomness can be a satellite or a source distributing quantum states or a supra-quantum attacker distributing local and non-local strategies (as we will see in chapter 5)

Assuming that Eve's error rate is lower than the honest parties, secret key agreement is surprisingly possible (unless Eve has noiseless access to the source's bits). We are going to describe the three phases of a secret key agreement protocol. As mentioned, Alice and Bob can start in a situation where the adversary has an advantage in terms of information retrieval. The objective of the first phase of advantage distillation is to generate an advantage over the opponent by exploiting the authenticity of the public channel. However, Alice and Bob do not generally share a mutual string after this phase. Hence a possibly interactive error correction phase called information reconciliation can be necessary. Finally, the resulting mutual but only partially secret string must be transformed into a shorter and highly secret key. This is privacy amplification. In the illustration 3.4, this process is shown graphically, with the relations between the amounts of information that Bob and Eve have of Alice's string.

In the following sections we develop each one of these phases, with special emphasis in advantage distillation that will be employed in the protocol described in chapter 5

3.2.1 Advantage distillation

The idea of advantage distillation is that Alice picks out several instances where she got the same bit and communicates the instances - but not the bit - to Bob. Bob replies yes only if it happens that for all these instances he also has the same bit value. For large error rates this is unlikely, but when it happens there is a large chance that both have the same bit. Eve can't influence the choice of the instances. All she can do is to use a majority vote for the cases accepted by Bob. The probability that Eve makes an error can be much larger than the probability that Bob makes an error (i.e. that all his instances are wrong), even if Eve's initial information is larger than Bob's.

Let's assume the scenario described in figure 3.3 with error probabilities $0 < \epsilon_A, \epsilon_B < \frac{1}{2}$ and $0 < \epsilon_E < \min\{\epsilon_A, \epsilon_B\}$. In this case, the adversary has an initial advantage over the trusted parties in terms of the error probabilities. Considering N independent realizations of the random variables, we have

$$\begin{aligned} I(A^N;B^N) &= \textstyle\sum_{i=1}^N I(A_i;B_i) = & N(1-h(\epsilon_A(1-\epsilon_B)+(1-\epsilon_A)\epsilon_B)) \\ I(A^N;E^N) &= \textstyle\sum_{i=1}^N I(A_i;E_i) = & N(1-h(\epsilon_A(1-\epsilon_E)+(1-\epsilon_A)\epsilon_E)) \\ I(B^N;E^N) &= \textstyle\sum_{i=1}^N I(B_i;E_i) = & N(1-h(\epsilon_B(1-\epsilon_E)+(1-\epsilon_B)\epsilon_E)) \end{aligned}$$

[1]This scenario is described on Maurer's paper [12]

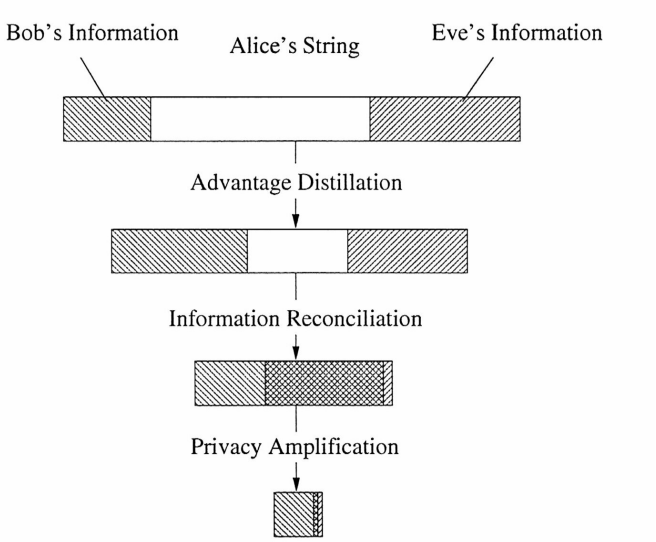


Figure 3.4: Phases of a secret-key agreement protocol: A first phase, advantage distillation, to generate an advantage over the opponent by exploiting the authenticity of the public channel. A second phase, a (possibly interactive) error-correction phase, information reconciliation, is required. Finally, the resulting mutual but only partially-secret string must be transformed into a (shorter) highly- secret string. This final phase is called privacy amplification.

and we have

$$I(A^N;B^N) < \min\{I(A^N;E^N);I(B^N;E^N)\} \tag{3.8}$$

The basic idea of the advantage distillation phase is that Alice and Bob in this setting use the public communication authenticated channel to exchange information about their bits in an insecure way to identify bits that are correct with a higher probability than others. the effect of advantage distillation can be seen as transforming a situation $\mathrm{I}(A^N;B^N) < \mathrm{I}(B^N;E^N)$ into a situation $\mathrm{I}(\tilde{A};\tilde{B}) > \mathrm{I}(\tilde{B};\tilde{E})$ where $\tilde{A},\tilde{B}$ represent the result of the processing made by the parties.

Advantage distillation: example 1

In order to introduce the application of advantage distillation to the protocol in chapter 5, let's develop the example from above to obtain the bounds necessary for advantage distillation to work.

Let's suppose that the scenario from figure 3.3 has the following properties:

- There is symmetry in the errors made by Alice and Bob ($\epsilon_A = \epsilon_B$) and thus $P_{A|R}(0,0) = P_{B|R}(0,0) = 1-\epsilon_A$. In the case of Eve, we have $P_{E|R}(0,0) = 1-\epsilon_E$
- Alice sends a randomly chosen bit C and computes and sends over the public channel $A^N = [C \oplus A_1, \ldots, C \oplus A_N]$, with N fixed.
- Bob computes $B^N = [(C \oplus A_1) \oplus B_1, \ldots, (C \oplus A_N) \oplus B_N]$ and accepts if this is equal to either $[0,0,\ldots,0]$ or $[1,1,\ldots,1]$

Bob accepts only if Alice and Bob have the same bit in all positions $A^N = B^N$ or opposite values in each position $A^N = B^N \oplus [1,1,\ldots,1]$. This way, they improve their position compared to the opponent's by accepting only in situation of highly reliable transmission. In Eve's case, she can compute $(C^N \oplus A^N) \oplus E^N$ and guess $C = 0$ if at least half of the bits in the string are 0 or 1 otherwise. The idea behind this process is that, for all possible choices of ϵ_A and ϵ_E, even if Eve's error is lower than Alice's and Bob's, Bob's error probability β_N about a bit sent by Alice decreases for $N \to \infty$ asymptotically faster than Eve's average error probability γ_N when Eve's uses the strategy described, which is optimal.

Proposition 3.2 *For the protocol described, there exist real-valued positive constants b and c with $b < c$ such that $\beta_N \leq b^N$ and $\gamma_N \geq c^N$ for sufficiently large N*

Proof: In order to prove the statement we must use

$$\binom{N}{N/2} = \frac{N!}{((N/2)!)^2} \geq \frac{1}{2}\frac{N^N\sqrt{2\pi N}e^N}{e^N(N/2)^N\pi N} = \frac{1}{\sqrt{2\pi N}}2^N \tag{3.9}$$

which is derived from Stirling's formula $lim_{n\to\infty}\frac{n!}{(n/e)^n\sqrt{2\pi n}} = 1$. Let $\alpha_{rs}(r,s \in 0,1)$ be the probability that the single bit 0 is received by Bob as r and by Eve as s Then we have,

$$\begin{aligned} \alpha_{00} = &\quad (1-\epsilon_A)^2(1-\epsilon_E) + \epsilon_A^2\epsilon_E \\ \alpha_{01} = &\quad (1-\epsilon_A)^2(\epsilon_E + \epsilon_A^2(1-\epsilon_E) \\ \alpha_{10} = &\qquad \alpha_{11} = (1-\epsilon_A)\epsilon_A \end{aligned} \tag{3.10}$$

Let $p_{\alpha,N}$ be the probability that Bob accepts the message sent by Alice. Assuming that N is even, then

$$\beta_N = \quad \tfrac{1}{p_{\alpha,N}}(\alpha_{10}+\alpha_{11})^N \quad = \frac{1}{p_{\alpha,N}}(2\epsilon_A - 2\epsilon_A^2)^N \tag{3.11}$$

$$\gamma_N \geq \quad \tfrac{1}{2}\tfrac{1}{p_{\alpha,N}}\binom{N}{N/2}\alpha_{00}^{N/2}\alpha_{01}^{N/2} \tag{3.12}$$

The last expression is half of the probability that Bob receives the correct codeword and that Eve receives the same number of 0's and 1's given that Bob accepts. It gives a lower bound on Eve's average error probability when she guesses C for all possible strategies because in this case, she does not obtain information about the bit C and half of the guesses will be incorrect. Employing the previous result, we have

$$\gamma_N \geq \quad \tfrac{1}{2}\tfrac{1}{p_{\alpha,N}}\tfrac{1}{\sqrt{2\pi N}}2^N\sqrt{\alpha_{00}\alpha_{01}}^N = \tfrac{K(2\sqrt{\alpha_{00}\alpha_{01}})^N}{\sqrt{N}p_{\alpha,N}}$$

for some constant K and sufficiently large N. For $0 < \epsilon_E \leq \frac{1}{2}$ we have

$$\sqrt{\alpha_{00}\alpha_{01}} = \sqrt{(1-2\epsilon_A+\epsilon_A^2-\epsilon_E+2\epsilon_A\epsilon_E)(\epsilon_A^2-2\epsilon_A\epsilon_E+\epsilon_E)} > \epsilon_A - \epsilon_A^2 \tag{3.13}$$

From

$$(1-2\epsilon_A+2\epsilon_A^2)^N \leq p_{\alpha,N} = (1-2\epsilon_A+2\epsilon_A^2)^N + (2\epsilon_A-2\epsilon_A^2)^N < 2(1-2\epsilon_A+2\epsilon_A^2)^N \tag{3.14}$$

We conclude that $\beta_N \leq b^N$ and $\gamma_N \geq c^N$ for N large enough and for

$$\begin{aligned} b = &\quad \tfrac{2\epsilon_A-2\epsilon_A^2}{1-2\epsilon_A+2\epsilon_A^2} \\ c = &\quad \tfrac{2\sqrt{\alpha_{00}\alpha_{01}}}{1-2\epsilon_A+2\epsilon_A^2} - \delta \end{aligned}$$

where δ can be arbitrarily small for large N. This concludes that $c > b$ □

The fact that Eve's error probability is greater than Bob's does not automatically imply that a key rate can be generated. We will see in the next section for this example the condition fulfilled to generate a secret key.

Advantage distillation: example 2

In a different setup where the information is not coming from a source but instead shared by Alice and Bob, we will not use ϵ_A and ϵ_B but instead e_{AB} as the error probability between Alice and Bob. This situation is the one that we will employ in chapter 5. For this version of advantage distillation, Alice reveals N instances such that her N bits are equal: $a_{i_1} = ... = a_{i_N} = \alpha$. Bob looks at the same instances, and announces whether his bits are also all equal. If indeed $b_{i_1} = ... = b_{i_N} = \beta$,

which happens with probability $(1-e_{AB})^N + e_{AB}^N$, Alice and Bob keep one instance; otherwise, they discard all the N bits. Bob's error on Alice's symbols becomes

$$\tilde{e}_{AB} = \frac{e_{AB}^N}{(1-e_{AB})^N + e_{AB}^N} \approx \left(\frac{e_{AB}}{1-e_{AB}}\right)^N . \tag{3.15}$$

Notice that $\tilde{e}_{AB} \to 0$ in the limit $N \to \infty$: this means that $\alpha = \beta$ almost always, for N sufficiently large. This remark is used to estimate Eve's probability of error.

As in the previous example, one finds that Eve's error on Bob's symbols goes as

$$\tilde{e}_E \gtrsim C\,(f(e_{AB}))^N \tag{3.16}$$

with $f(.)$ some function which depends on the probability distribution under study. Now, as long as the condition

$$f(e_{AB}) > \frac{e_{AB}}{1-e_{AB}} \tag{3.17}$$

is fulfilled, Eve's error at the end of advantage distillation is exponentially larger than Bob's for increasing N: there exists always a finite value of N such that Eve's error becomes larger than Bob's. The bound on the tolerable error after advantage distillation is then computed by solving eq. 3.17.

3.2.2 Privacy amplification and information reconciliation

After advantage distillation, Alice and Bob have computed strings S_A and S_B respectively and Bob has more information about Alice's string than the adversary. However, Eve also has information about the content of the strings and the goal of the key-agreement protocol is that Alice's and Bob's strings are equal and secure with high probability.

Information reconciliation is error-correction conducted over a public channel, which reconciles errors between S_A and S_B to obtain a shared bit string S while divulging as little as possible to Eve. After this procedure, suppose Eve has obtained a random variable E which is partially correlated with S.

Since Alice and Bob must share a common string after information reconciliation, this leads to a lower bound on the amount of error-correction information C that must be exchanged. Bob has to know S_A exactly with high probability after this step, given S_B and C, therefore

$$0 \approx H(S_A|S_B, C) \geq H(S_A|S_B) - H(C)$$

And hence,

$$H(C) \gtrsim H(S_A|S_B))$$

And the uncertainty about S_A that Eve has can as well be reduced by $H(C)$ when she learns C.
A good protocol for information reconciliation should minimize the amount of information leaked to the adversary and be efficient.

The next step is privacy amplification: shrink a partially-secure string S of length n to a highly-secret string S' by public discussion.
Proposed protocols for privacy amplification consists on Alice choosing a suitable number of random bits and sending them publicly to Bob. From these bits, Alice and Bob then compute a compressed string. One way to compute the string S' uses the class of universal hash functions $\mathcal{G}$, which map the set of n-bit strings $\mathcal{A}$ to the set of r-bit strings $\mathcal{B}$, such that for any distinct S_1, $S_2 \in \mathcal{A}$, when g is chosen uniformly at random from $\mathcal{G}$, then the probability that $g(a1) = g(a2)$ is at most $1/|\mathcal{B}|$. The central question is thus how long the virtually secure string S' can be depending on the hashing technique and on the type and amount of Eve's knowledge about S. We can express that Eve has some information about S as given Eve's entire knowledge $U = u$ about S, the random variable S is not uniformly distributed and thus $H(S|U = u) < n$. Eve has then $n - H(S|U = u)$ bits of information about S. The resulting string must satisfy $H(S'|C, U = u) \approx r$ where r is the length

of S' and C is the communication between the parties. Therefore, if Eve has t bits of information about S, the length of the resulting string S' can be $n-t$. This fact was shown to be correct in the case if Eve has deterministic information about S. However if Eve information is not deterministic, it is not true in general $n-t$ bits can be extracted when Eve has t bits of Shannon information. In the case of universal hash functions it was shown that the Renyi entropy[2] $H_2(S|U=0)$ is the right measure for the adversary's information. We don't detail anymore these phases of secret key agreement because they are out of the scope of the work done for the analyzed protocol.

3.3 Key-Rate bound for secrecy extraction

The knowledge of the exact relationship between the secret key rate $S(A;B \parallel E)$ and the joint probability distribution P_{ABE} is a challenging problem, to determine the value of $S(A;B \parallel E)$ or at least decide whether this rate is non-zero.

We will describe the Csiszár-Körner (CK) bound for secret key rate that will be useful to obtain bounds in the protocol presented. We assume that the parties are connected with conditional output distribution $P_{BE|A}$. The ability of generating mutual mutual secret information is quantified as follows

Definition 3.6 *Consider a channel characterized by the conditional joint distribution $P_{BE|A}$. The secrecy capacity*

$$C_S(P_{BE|A})$$

of the channel is the maximal real number $R \geq 0$ such that for every $\epsilon > 0$ and for $K = \lceil (R-\epsilon)N \rceil$, there exist a possibly probabilistic (additionally depending on some random bits) encoding function

$$\xi : 0,1^K \to \mathcal{A}^N$$

together with a decoding function

$$\nu : \mathcal{B}^N \to 0,1^K$$

such that if S is uniformly distributed over $0,1^K$, we have $A^N = \xi(S)$, $S' = \nu(B^N)$ and $P_{B^N|A^N} = P_{B|A}^N$ (being $P_{B|A}^N$ the marginal of $P_{BE|A}$)

$$P\left[S \neq S'\right] < \epsilon$$

and

$$\frac{1}{K} H(S|E^N) > 1-\epsilon \tag{3.18}$$

The result proved by Csiszár-Körner in [32] is that

$$C_S(P_{BE|A}) \geq \max_{P_A} (I(A;B) - I(A;E)) \tag{3.19}$$

This result will be employed in chapter 5 in the following way: in the case that Eve knows more about Alice's symbol than about Bob's, the C-K bound can be expressed as

$$R_{CK} = \sup_{(B',T)\leftarrow B} \left[I(A;B') - I(B';E)\right] = \sup_{(B',T)\leftarrow B} \left[H(B'|E,T) - H(B'|A,T)\right] \tag{3.20}$$

employing $I(X;Y) = H(X) - H(X|Y)$ and where $B \to (B',T)$ is called pre-processing: from his initial data B, Bob obtains some processed data B' that he does not reveal, and some other processed data T that are broadcasted on a public channel. For classical distributions, bitwise pre-processing is already optimal.

We define next the secret key rate for a secret-key agreement protocol.

[2]It is defined from the collision probability of a random variable X, $P_C(X) = \sum_{x\in\mathcal{X}} P_X(x)^2$, as $H_2(X) = -\log(P_C(X))$

Definition 3.7 *A (P_{ABE}, r, ϵ) protocol for secret key agreement is defined by two phases:*

1. *A communication phase where Alice and Bob exchange messages $C_1, C_2, \ldots$ over the public channel. We assume that Alice sends odds messages and Bob even messages. A message sent at some point in the protocol can only depend on the sender's knowledge at that moment and possibly some locally generated random bits R_A or R_B. We have for odd i*

$$H(C_i|A, C_1, \ldots, C_{i-1}, R_A) = 0 \tag{3.21}$$

and for even i

$$H(C_i|B, C_1, \ldots, C_{i-1}, R_B) = 0 \tag{3.22}$$

If t messages are shared during this communication phase, let

$$C = [C_1, C_2 \ldots, C_t]$$

be the collection of all messages.

2. *A key generation phase where Alice and Bob compute keys S and S' respectively. Their goal is to minimize the probability that $S \neq S'$ and the information that the attacker obtains about S. We can write this as follows:*

$$\begin{aligned} H(S|CAR_A) &= 0, \\ H(S'|CBR_B) &= 0, \\ H(S) &\geq r, \\ P[S \neq S'] &\leq \epsilon, \\ I(S; EC) &\leq \epsilon \end{aligned}$$

for some independent random bit strings R_A and R_B.

We define the secret key rate as follows

Definition 3.8 *The secret key rate $S(A; B \parallel E)$ is the least upper bound of the set of real numbers $R \geq 0$ such that for every large $N = N(\epsilon)$, there exists a*

$$(P^N_{ABE}, (R - \epsilon)N, \epsilon)$$

protocol[3] *for secret key agreement such that it satisfies that Alice and Bob can compute strings S and S' respectively such that*

$$\begin{aligned} H(S)/N &\geq R - \epsilon, \\ P[S \neq S'] &\leq \epsilon, \\ I(S; E^N C) &\leq \epsilon \\ H(S) &\geq log|S| - \epsilon, \end{aligned}$$

The value of the secret key rate is an intrinsic property of the joint distribution of three random variables. In [12] the following bound was proved that states that is impossible to generate a greater amount of key from A and B than the mutual information between A and B (given E)

$$S(A; B \parallel E) \leq \min I(A; B), I(A; B|E) \tag{3.23}$$

The following result states a nontrivial lower bound on the secret key rate. If either Eve has less information on B than Alice or symmetrically less information on A than Bob, then this difference in information can be exploited.

Theorem 3.2 *For every distribution P_{ABE},*

$$S(A; B \parallel E) \geq \max\{I(A; B) - I(A; E), I(A; B) - I(B; E)\} \tag{3.24}$$

[3] We are employing a large number of repetitions of the same random experiment and $P^N_{ABE} = \Pi^N_{i=1} P_{A_i, B_i, E_i}$

The proof of the previous theorem is referenced in [14].

At last we will prove the result previously stated in the advantage distillation example 1, that the bounds expressed lead to the extraction of a secret key. We provide this result to give an example of how to obtain the bound calculation that is useful for the non-signaling protocol analysis in chapter 5.

Proposition 3.3 *In the conditions described in advantage distillation example 1, let C the random bit selected by Alice and M the message generated from A^N and C such that Bob accepts publicly and computes a bit C' with some probability $P_{\alpha,N}$, such that $P[C \neq C'] \leq b^N$ for some $b \geq 0$. If in addition given that Bob accepts, for every strategy for guessing C when given M and E^N, Eve's average error probability γ_N is at least c^N for some $c > b$ and sufficiently large N, then $S(A;B \parallel E) \geq 0$*

Proof: From the previous theorem, it must be proved that for sufficiently large N

$$I(\hat{A};\hat{B}) - I(\hat{A};\hat{E}) > 0$$

with $\hat{A}$ and $\hat{B}$ being random variables constructed after the interaction and $\hat{E} = \left[E^N, V\right]$ where V is the collection of all messages sent over the public channel. Defining $\hat{A} = C$ and $\hat{B} = C'$ if Bob accepts and $\hat{A} = \hat{B} =$ reject if Bob rejects. If Bob accepts, we have,

$$H(C|C') \quad \leq h(b^N) \leq 2b^N \log(1/b^N) = 2b^N N \log(1/b) < c^N$$

for N sufficiently large. The first inequality comes from Jensen's lemma and the second one is because for $p \leq 1/2$ we have $-p\log p \leq -(1-p)\log(1-p)$. The third inequality is obtained for a large N.

$$H(C|\hat{E}) \quad = \textstyle\sum_{\hat{e}\in\mathcal{E}\times\mathcal{V}} P_{\hat{E}}(\hat{e}) H(C|\hat{E}=\hat{e}) = E_{\hat{E}}(h(p_{\epsilon,\hat{E}})) \geq E_{\hat{E}}(p_{\epsilon,\hat{E}}) = \gamma_N \geq c^N \tag{3.25}$$

where $p_{\epsilon,\hat{E}}$ is the probability that Eve guesses incorrectly C with her optimal strategy given $\hat{E} = \hat{e}$. As $p_{\epsilon,\hat{E}} \leq 1/2$ then $h(p_{\epsilon,\hat{e}}) \geq p_{\epsilon,\hat{e}}$ for all $\hat{e}$. Given that Bob publicly rejects then $H(\hat{A}|\hat{B}) = H(\hat{A}|\hat{E}) = H(\hat{A}|V) = 0$. As the probability that Bob accepts some C is > 0, then $I(\hat{A};\hat{B}) - I(\hat{A};\hat{E}) = H(\hat{A}|\hat{E}) - H(\hat{A}|\hat{B}) > 0$ □

3.4 Intrinsic information

In this section, we describe a information measure called the intrinsic conditional information, that is an upper bound for the secret-key rate and useful for distinguishing between cases where secret-key agreement is possible from those where it is impossible. This section is based in the references [14] and [33].

Definition 3.9 *Let P_{XYZ} a discrete probability distribution. Then we can define the **intrinsic conditional mutual information between** X **and** Y **given** Z as*

$$\mathrm{I}(X;Y\downarrow Z) = \inf_{XY\to Z\to\bar{Z}} \{I(X;Y \mid \bar{Z}) : P_{XY\bar{Z}} = \sum_{z\in\mathcal{Z}} P_{XYZ} P_{\bar{Z}}\} \tag{3.26}$$

where the infimum is taken over all possible conditional distributions $P_{\bar{Z}|Z}$

This measure has the following properties

$$\begin{aligned} 0 \quad & \mathrm{I}(X;Y\downarrow Z) \leq I(X;Y) \geq \\ & \mathrm{I}(X;Y\downarrow Z) \leq I(X;Y|Z) \geq \\ & \mathrm{I}(X;Y\downarrow Z) \leq \mathrm{I}(X;Y\downarrow \bar{Z}) \geq \end{aligned}$$

where $\bar{Z}$ is generated by sending Z over an arbitrary channel (that is $XY \to Z \to \bar{Z}$ is a Markov chain [4]. The idea behind the measure is that giving the side information Z in some way destroys

[4]We recall that a sequence of random variables $X_1, X_2, \ldots, X_N$ is called a Markov chain, denoted by $X_1 \to X_2 \to \ldots \to X_N$ if for all $i > 1$, we have $P_{X_i|X_1X_2\ldots X_{i-1}} = P_{X_i|X_{i-1}}$).

all the information between X and Y but generates new conditional information that cannot be employed to generate a secret key. Instead, $\mathrm{I}(X;Y \downarrow Z)$ measure only the remaining conditional mutual information between X and Y but not the additional information brought in by Z.

The figure 3.5 represents the idea behind this measure. The particular $\bar{Z}$ represented is the one that minimizes the region and therefore $\mathrm{I}(X;Y \mid \bar{Z})$ and fulfills $\mathrm{I}(X;Y \downarrow \bar{Z})$.

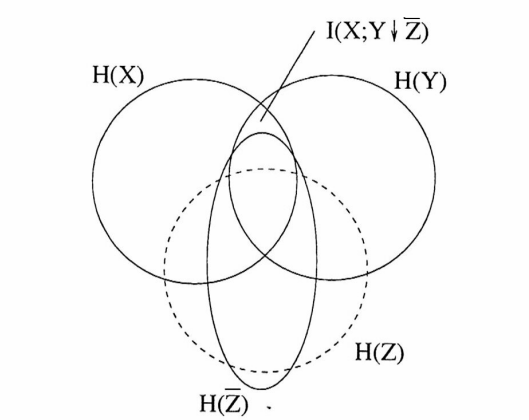


Figure 3.5: Representation of intrinsic information $\mathrm{I}(X;Y \downarrow Z)$

We now state two theorems that are important related to this measure. The first one gives the relationship with the secret-key rate, where we employ the notation typical for a secrecy extraction scenario and we write the intrinsic information as

$$I(A:B \downarrow E) = \min_{E \to \bar{E}} I(A:B|\bar{E}), \tag{3.27}$$

Theorem 3.3 *For arbitrary random variables A, B, and E representing the three parties in a secret key agreement scenario, we have*

$$S(A;B \parallel E) \leq \mathrm{I}(A;B \downarrow E) \tag{3.28}$$

This theorem implies that secret-key agreement can be possible only if $\mathrm{I}(A;B \downarrow E) > 0$. Then it is a necessary condition to obtain secrecy (whether it is sufficient will be discussed at the end of the section). Therefore computing this measure can help in discarding possible candidate protocols, but the intrinsic information is hard to compute given that the minimization is calculated over all discrete random variables $\bar{E}$ such that $AB \to E \to \bar{E}$ is a Markov chain, i.e. in all discrete conditional distributions or discrete channels $P_{\bar{E}|E}$.

The following theorem proved in [33] offers a simplification to calculate this measure in the case of finite alphabets. In the theorem it is stated that the minimization in Eq. (3.27) can be restricted to variables $\bar{E}$ of the same size as the original one, E. This allows a numerical approach to this problem.

Theorem 3.4 *If the range $\mathcal{E}$ of E is finite, then there exists a finite random variable having the same range $\mathcal{E}$, such that $AB \to E \to \bar{E}$ is a Markov chain and*

$$\mathrm{I}(A;B \downarrow E) = \mathrm{I}(A;B \mid \bar{E})$$

and then

$$\mathrm{I}(A;B \downarrow E) = \min_{\bar{E}} \mathrm{I}(A;B \mid \bar{E})$$

The intrinsic information can be understood as a witness of secret correlations in $P(a,b,e)$. Indeed, a probability distribution can be established by local operations and public communication

if, and only if, its intrinsic information is zero [34]. It is then clear why the positivity of the intrinsic information is a necessary condition for positive secret-key rate. Whether it is sufficient is at present unknown: strong support has been given to the existence of probability distributions such that the secret key rate $R = 0$ and $\mathrm{I}(A;B \downarrow E) > 0$. These would constitute examples of probability distributions containing bound information [35], that is non-distillable secret correlations. The existence of bound information has been proven in a multipartite scenario consisting of $N > 2$ honest parties and the eavesdropper [36]. However, it (still) remains as an open problem for the more standard bipartite scenario.

In this chapter we have provided the tools from information-theoretic secret key agreement to analyze the protocol in chapter 5.

Chapter 4

Non-locality and non-signaling

This chapter provides the framework developed in [15] (and previous papers) to study non-locality and non-signaling in a more general framework than Quantum Mechanics. It is a theoretical framework for a physical theory that do not allow superluminal signaling and predict the violation of Bell inequalities, i.e. a non-signaling non-local theory.

The constancy of the speed of light in any reference frame implies that no signal carrying information can propagate faster than light (no signaling principle).
The existence of correlations between spacelike separated events that violate Bell inequalities cannot contradict the no signaling principle and such correlations are called non-local. The idea to develop a physical theory that is consistent with the two principles can be very useful to understand Quantum Mechanics from the outside and gain more knowledge about the paper of the combination of these 2 principles in quantum mechanics. From an information-theoretical point of view, it is worth looking at a framework more general than Quantum Mechanics to analyze the use of nonlocal correlations as an information resource. The aim of the present chapter is to provide the theoretical foundations for the Secret Key Distribution protocol developed in the next chapter.

The kind of experiment-setup we are considering onwards is sketched in Fig. 4.1. A source sends out two particle to two distant locations. In each location, a user chooses a possible measurement (x, y) and registers the outcome (a, b). The procedure is repeated a large number of times.

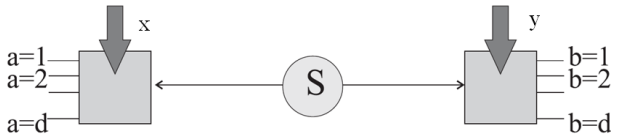


Figure 4.1: The setup to be kept in mind for Bell-type experiments: a source S distributed two quantum systems to separated locations. On each location, the physicist is free to choose which measurement to perform (x, y); as a result, they obtain outcomes a, b. Of course, this setup can be generalized to more parties, or to the case where the number of outcomes is different between the parties.

Later, the two users come together, compare their results and derive the probability distribution

$$P(a, b|x, y)\,. \tag{4.1}$$

This probability distribution, here written as a conditional probability, can also be written as $P_{xy}(a, b)$. This last notation can be employed to stress that the origin of the statistics on (a, b) is the quantum randomness (or more nonlocal, as we will see in the following pages) we want to query, while the statistics on (x, y) are of a very different nature (just the choice of the users on how often to perform each measurement).

By defining measurement in this way, we are in some way generalizing some aspects that were introduced in chapter 2: the correlations that arise in entangled states between two parties can be in this way studied independently of the particularities of such states as these particularities are contained in the probability distributions obtained.

A few crucial remarks:

(i) We can say that there is a conceptual distinction between the users and the physicists who have constructed the whole setup. To perform the task, the users do not even have to know what state the source has prepared, what physical systems are sent, which measurements are being performed. For the users, the measurement is just a knob they can freely set at any position; x and y refer to the label of the scale in their instruments. As in a secret key agreement protocol, the names of Alice and Bob will be always referring to the users. This assumption of lack of knowledge of the users of such a device is necessary to derive its use in cryptographic applications[1].

(ii) The number of possible outcomes a and b plays an important role, but the labeling of the outcomes does not matter: the probability distribution $P(a,b)$ fully characterizes the process of measuring x and y, irrespective of whether the outcomes are real numbers, multiples of $\hbar$, complex numbers, colors... Because of this, we are free to choose the most convenient labeling without loss of generality. For instance, if the number of outcomes is two on both sides, then the correlation coefficient is defined as

$$C_{xy} \quad = \quad P(a=b|x,y) - P(a\neq b|x,y)\,. \tag{4.2}$$

This definition is unambiguous and always valid. If in addition, one chooses the labeling $a,b\in\{-1,+1\}$, then one obtains the relation $C_{xy}=\langle ab\rangle_{xy}$ the average of the product of the outcomes.

4.1 Classical mechanisms for correlations

In the setup defined are dealing with a family of probability distributions: Alice picks up measurement x, Bob picks up measurement y, and the outcomes a,b are guaranteed to be distributed according to the probability distribution $P(a,b|x,y)$. Now, in general, $P(a,b|x,y)\neq P(a|x)P(b|y)$ where $P(a|x)=\sum_b P(a,b|x,y)$ and $P(b|y)=\sum_a P(a,b|x,y)$ are the marginals of the probability distribution. In other words, random variables distributed according to $P(a,b|x,y)$ are correlated. The question that we are going to study is: can these correlations be ascribed to a classical mechanism? The question seems quite open, but it becomes more concrete once one takes into account that there are only two classical mechanisms for distributing correlations.

- The first and most obvious is communication: for instance, the information about Alice's choice of measurement x is available Bob when he measures his particle. This mechanism can be checked by arranging what is known as space-like separation: loosely speaking, if the two particles reach the measurement devices at the same time and the choice of measurement is done at the very last moment, a signal informing a particle about what is happening in the other location should travel faster than light. If the correlations persist in this configuration, the mechanism of communication can't explain such correlations.

- The second mechanism consists in using pre-established strategies: each particle might have left the source with a set of instructions, specifying how it should behave in any measurement. Interestingly, this mechanism can explain the behavior of the singlet state in the cases where Alice and Bob choose to perform the same measurement ($P_{MM}=P(\cdot|M,M)$). Indeed, assume that the particles are carrying lists of pre-determined results $\lambda_x=\{a_M\}_M$ and $\lambda_y=\{b_M\}_M$ such that $a_M=-b_M$ for each measurement M: one always gets $P_{MM}(a\neq b)=1$, and the local randomness can be easily accounted for by varying the lists at each run. However, it is the milestone result by John Bell in 1964 to prove that the whole family of probabilities predicted by quantum physics cannot be reproduced with pre-established strategies [10], as we described in chapter 2.

Let's detail more properties about pre-established strategies that will be useful for the purpose of our cryptographic protocol.

[1] As we will explore in the next chapter, this lack of knowledge is, from a cryptographic point of view, a necessary assumption as the devices could be provided by the eavesdropper Eve. Additionally, the prerequisite "must know experimental quantum physics before using" is not the best way to commercialize quantum key distribution systems.

Pre-established strategies ("Local variables")

By definition, a pre-established strategy is some hypothetical information λ that the particles carry out with themselves from the source. Each particle is supposed to produce its outcome taking into account only this λ and the measurement to which it is submitted. In other words, for given λ, the two random processes are supposed to be independent: $P(a,b|x,y,\lambda) = P(a|x,\lambda)P(b|y,\lambda)$. The only freedom left is the possibility of changing the information λ at each run. If λ is drawn from a distribution $\rho(\lambda)$, the observed probability distribution will be[2]

$$P(a,b|x,y) \quad = \quad \int d\lambda\, \rho(\lambda)\, P(a|x,\lambda)P(b|y,\lambda)\,. \tag{4.3}$$

Here comes an important mathematical characterization:

Theorem 4.1 *$P(a,b|x,y)$ can be obtained by pre-established strategies if and only if it can be written as a convex sum of local deterministic strategies.*

Proof: A *deterministic strategy* is a strategy in which, for each possible measurement, the result is determined. A *local* deterministic strategy is defined by $P(a|x,\lambda) = \delta_{a=f(x,\lambda)}$ and $P(b|y,\lambda) = \delta_{b=g(y,\lambda)}$. The "if" implication is therefore trivial.

The "only if" implication stems from the fact that any classical random process can be mathematically decomposed as a convex sum of deterministic processes. We are going to construct a deterministic model that gives the same statistics as the initial stochastic model. Let's label the possible values of a as $\{1,2,\ldots,m_A\}$. We define the cumulative distribution as $\mathcal{F}(a) = \sum_{\alpha\leq a} P(\alpha|x,\lambda)$. Adding a new local parameter μ_A, distributed as a uniform distribution between 0 and 1, then output a according to the following deterministic rule:

$$P_d(a|x,\lambda,\mu) = \begin{cases} 1 & \text{if } \mathcal{F}(a-1) \leq \mu_A < \mathcal{F}(a) \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

As μ is drawn with uniform distribution, the original stochastic distribution is recovered

$$\int_0^1 d\mu P_d(a|x,\lambda,\mu) = \int_{\mathcal{F}(a-1)}^{\mathcal{F}(a)} d\mu = P(a|x,\lambda)$$

Equation 4.3 can be rewritten as

$$P(a,b|x,y) = \int d\lambda\rho(\lambda) \int_0^1 d\mu_A \int_0^1 d\mu_B P_d(a|x,\lambda,\mu_A)P_d(b|y,\lambda,\mu_B) \tag{4.5}$$

which is the sum of local deterministic strategies, for the enlarged variable $\lambda' \equiv (\lambda,\mu_A,\mu_B)$ distributed as $\rho'(\lambda')d\lambda' = \rho(\lambda)d\lambda d\mu_A d\mu_B$.

In other words, for each λ, there exist an additional random variable $\mu = \mu(\lambda)$ with distribution $\rho'(\mu)$ such that $P(a|x,\lambda) = \int d\mu\rho'(\mu)\delta_{a=f(x,\lambda,\mu)}$. One can therefore just "enlarge" the definition of the local variable to $\lambda' = (\lambda,\mu(\lambda))$.

□

This theorem is a purely mathematical result: no restriction on the family of pre-established strategies under study is being made. It is useful because it allows deriving results by arguing with deterministic strategies, that are very easy to handle; if the result is stable under convex combination, it is automatically guaranteed to hold for all possible pre-established strategies, deterministic or not.

[2]This can also be expressed equivalently in terms of sums in the discrete case: $P(a,b|x,y) = \sum_\lambda P(\lambda)P(a|x,\lambda)P(b|y,\lambda)$

History and terminology

Historically, λ was called "local hidden variable". While the expression "local variable" is ultimately convenient, the adjective "hidden" is definitely superfluous and even misleading: quantum physics is at odds with local variables, irrespective of whether they are supposed to be hidden or not. For instance, the local variable may be a description of the total quantum state. Much more recently, in the interaction between physicists and computer scientists, the name of "shared randomness" has also become fashionable to denote pre-established strategies. We will use the term "non-locality" and talk about to "non-local correlations" as a shortcut for "probability distributions that cannot be reproduced by pre-established agreement".

CHSH inequality, revisited

We have already stated the CHSH inequality as a Bell inequality whose violation exposes the nonlocal nature of quantum physics. Let's derive its expression attending to the previous notation. The definition of this inequality is based in the fact that Alice and Bob can make only two possible measurements and the outcomes are binary. Let's label the measurements options as x,x' for Alice and y, y' for Bob. The outcomes are labeled as $a, b \in +1, -1$.
We consider a local deterministic strategy λ_L, which is just a list $\lambda_L = (a, a', b, b')$ specifying the two outcomes of Alice and Bob. The number $S(\lambda_L) = (a + a')b + (a - a')b'$ is defined. Substituting the different possible values of a and b, by inspection it is clear that $-2 \leq S(\lambda_L) \leq 2$. If we take a convex combination of local deterministic strategies, with distribution ρ, the value $\langle S \rangle = \int d\lambda_L \rho(\lambda_D) S(\lambda_D)$ must lie between -2 and 2.
Since we have:

$$\langle S \rangle = \langle ab \rangle + \langle a'b \rangle + \langle ab' \rangle - \langle a'b' \rangle$$

and with the labeling employed we have that the averaged products are the correlation coefficients $C_{xy} = \langle ab \rangle$ and $C_{x'y'} = \langle a'b' \rangle$, therefore

$$| \langle S \rangle | = | C_{xy} + C_{x'y} + C_{xy'} - C_{x'y'} | \tag{4.6}$$

This is the CHSH inequality. By the theorem 4.1, it holds for all pre-established strategies. When we employ the outcomes labeled as +1 and -1, the derivation of this inequality is simple. But the inequality is independent of this choice. For the general definition of the correlation coefficients 4.2, the inequality still holds.

4.2 The mathematics of no-signaling

The definition of quantum physics is a description of its formalism. There is a long list of "typically quantum features": intrinsic randomness, incompatible measurements and uncertainty relations, no-cloning, non-local correlations... Popescu and Rohrlich ([13]) asked the following question for non-locality without signaling: does quantum physics define the set of all probability distributions that are possibly non-local but are compatible with the non-signaling condition? The answer is: NO. Consider the following probability distribution for binary input $x, y \in \{0, 1\}$ and binary output $a, b \in \{0, 1\}$

$$P(a \oplus b = xy | x, y) = 1, P(a|x, y) = P(b|x, y) = \frac{1}{2} \tag{4.7}$$

Explicitly, each of the probabilities can be expressed as

$$\mathbf{P}_{PR} \quad : P(a, b|0, 0) = P(a, b|0, 1) = P(a, b|1, 0) = \tfrac{1}{2}\delta_{a=0,b=0} + \tfrac{1}{2}\delta_{a=1,b=1} \tag{4.8}$$

$$P(a, b|1, 1) = \tfrac{1}{2}\delta_{a=0,b=1} + \tfrac{1}{2}\delta_{a=1,b=0} \tag{4.9}$$

the distribution is non-signaling: the marginal of Alice does not depend on Bob's measurement (nor on Alice's in this case) and the same for Bob. The correlation says that for $(x, y) \in$

$\{(0,0),(0,1),(1,0)\}$ one has $a = b$ (perfect correlation) while for $(x,y) = \{(1,1)$ one has $a = b \oplus 1$ (perfect anti-correlation). Let's reobtain the CHSH inequality in order to calculate the value provided by this distribution.

If one consider the CHSH inequality, expressed as

$$|\langle S\rangle| = |C_{00} + C_{10} + C_{01} - C_{11}| \tag{4.10}$$

where we recall, $C_{xy} = P(a = b|x,y) - P(a \neq b|x,y)$ and labeling $a,b = 0$ as $+1$ and $a,b = 1$ as -1 (in order to make it compatible with the inequality). With the distribution proposed, one has:

$$C_{00} = 1\,, C_{01} = 1\,, C_{10} = 1\,, C_{11} = -1$$

Therefore if one evaluates the CHSH expression on this distribution, one finds $\langle S\rangle = 4$.

This innocent-looking probability distribution reaches the largest possible value of CHSH, while we saw in section 2.3 that quantum physics cannot go beyond the Tsirelson bound $\langle S\rangle = 2\sqrt{2}$. The hypothetical resource that would produce the probability distribution in 4.7 has been called *PR-box* (or non-local machine). After some time remaining in the shadows, it was hoped initially that the PR-box could be a building block for all non-local distributions. This was shattered as it was shown that the PR-box was not the only non-signaling resource outside of quantum physics.
This chapter provides the formal framework to study non-signaling distributions. It will be shown with an example the structure of such a distribution and afterwards it will be generalized. But to form the intuition of what we're talking about when we're talking about non-signaling non-local correlations, let's have a playful interlude:

4.2.1 3 games

Let's consider 3 games[3] in which two players, Alice and Bob, after learning the rules are sent to two different locations. There, each of them receives an external input and has to react. The game is won if the reactions are in agreement with the rules fixed at the beginning of the game.

Game 1: *each of the players receives a stock to invest in. The game is won if, whenever the two players receive the same stock, they produce the same answer (that is, they either both accept it or both reject it).*

This game is easy to win: Alice and Bob can agree in advance that whatever stock they receive, they will accept it. This strategy is boring and they can refine it a little more, as accept in even numbers and reject in odd numbers or establish a more elaborate strategy (as accept whenever [the first letter is between A-F] AND [the stock is from Japan], reject otherwise).
These are pre-established strategies and there are uncountably many of them.

Game 2: *each of the players receives a stock to invest in. The game is won if the two players produce the same answer only when they receive the same stock, otherwise they produce different answers.*

If the set of possible inputs consists only on two stocks, the game can still be won by pre-established strategies: They agree previously which stock accept and which reject. If there are more than two possible inputs, however, pre-established strategies cannot win the game with certainty.
For example, the case with 3 stocks and two possible outcomes, there will be uncertainty in 2 of the stocks. In this case, no local deterministic strategy can win the game.

To win this game, one needs in the classical world to use communication as a resource. But communication is an exaggerated resource, as with communication one can say that every game can be won. However, with what we know about the quantum world, there can be strategies that can make winning the game without communication (non local strategies).

Game 3: *each of the players receives a stock to invest in. The game is won under the condition: Alice invests in its stock if and only if the stock received by Bob is from Spain.*

Game 3 is more extreme than Game 2 because Alice must learn specific information about the input received by Bob, while in Game 2, winning criteria involves relations between Alice's and

[3]This is based in a course by Prof. V.Scarani in which this chapter is in part based but I preferred to change the scope of the game from papers to stocks, just for fun.

Bob's inputs.
Communication is required to win Game 3 but one might hope to win Game 2 without communication, only with no-signaling resources.

4.2.2 Formalization of no signaling principle and Bell-type experiments

The situation that one must keep in mind is a Bell-type experiment, that is, the measurement of non locality between two parties, Alice and Bob. Let's formalize the situation under study:

Alice and Bob receive several pairs of entangled quantum particles. On each particle, Alice performs the measurement x randomly drawn from a finite set of m_A possibilities; as a result, she obtains the output a out of a discrete set containing n_A symbols. Independently from Alice, Bob performs the measurement y randomly drawn from a finite set of m_B possibilities; as a result, he obtains the output b out of a discrete set containing n_B symbols. The results of such an experiment are characterized by the family of probabilities

$$P(a,b,x,y) \quad = \quad P(a,b|x,y)\,P(x)\,P(y)\,. \tag{4.11}$$

There are $D = m_A\,m_B\,n_A\,n_B$ such numbers, so each experiment can be described by a point in a D-dimensional space; more precisely, in a region of such a space, bounded by the conditions that probabilities must be positive and sum up to one. By imposing further restrictions on the possible probability distributions, the region of possible experiments shrinks, thus adding non-trivial boundaries.

The three important restrictions to be considered in order to limit the structure of such region of probabilities are the following:

- The first restriction is the requirement that the probability distribution must be built without communication, only with shared randomness. These are the conditions described in section 4.1 The bounded region, which contains all probability distributions that can be obtained by shared randomness, forms a polytope, that is a convex set bounded by a finite number of hyperplanes ("facets"); therefore we refer to it as to the *local polytope*. The vertices of the local polytope are the points corresponding to *deterministic strategies*, that is, strategies in which $a = a(x)$ and $b = b(y)$ with probability one; that is, $P(a,b|x,y) = \delta_{a,a(x)}\delta_{b,b(y)}$. There are clearly $m_A^{n_A}\,m_B^{n_B}$ such strategies. The vertices are thus easily listed, but to find the facets given the vertices is a computationally hard task. The importance of finding the facets is pretty clear. If a point, representing an experiment, lies within the polytope, then there exists a strategy with shared randomness that produces the same probability distribution. If on the contrary a point lies outside the local polytope, then the experiment cannot be reproduced with shared randomness only. The facets of the local polytope are Bell's inequalities. We shall call *non-local region* the region which lies outside the local polytope.

- The second restriction is the requirement that the probability distribution must be obtained from measurements on quantum bipartite systems. The bounded region thus obtained shall be called the *quantum region*. It is not a polytope, since there is not a finite set of extremal points. It is a convex set if one really allows all possible measurements on all possible states in arbitrary-dimensional Hilbert space. The quantum region contains the local polytope, but is larger than it: measurement on quantum states can give rise to non-local correlations (Bell inequalities are violated).

- The third and more general restriction is the requirement that the probability distribution must not allow signaling from Alice to Bob or viceversa. The no-signaling requirement is fulfilled if and only if Alice's marginal distribution does not depend on Bob's choice of input, and viceversa: that is, the probability distributions must fulfill

$$\sum_b P(a,b|x,y) \quad = \quad P(a|x)\,, \tag{4.12}$$

$$\sum_a P(a,b|x,y) \quad = \quad P(b|y)\,. \tag{4.13}$$

  These conditions define again a polytope, the *no-signalling polytope*, which contains the quantum region. The deterministic strategies are still vertices for this polytope; to these, one must

add other vertices which represent, loosely speaking, purely non-local no-signalling strategies, as the PR-box.

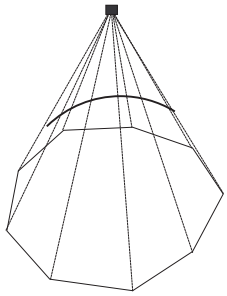


Figure 4.2: Figurative representation of the nonlocal polytope. The base represents the local polytope and the black box represents the nonlocal machine (or PR-box). The quantum boundary is schematically represented.

4.2.3 Local and no-signaling polytopes, case $d = 2$

Let's consider in first place the case $n_A = n_B = 2$ and $m_A = m_B = d = 2$ to introduce the no-signaling polytope and a useful notation that we will use for the case where $d > 2$. Let's suppose that Alice and Bob each have two inputs $x, y \in \{0, 1\}$ and two outputs $a, b \in \{0, 1\}$. What we want to characterize is the joint probability conditioned to the inputs $P(a, b \mid x, y)$. The first element to be studied is the dimensionality of the probability space: how many numbers are required to specify the four no-signaling probability distributions $P(a, b \mid x, y)$ completely? There are 16 probabilities, but since $\sum_{a,b} P(a, b \mid x, y) = 1$ for each x, y it reduces to 12 the number of probabilities.
A subsequent reduction can be made to only eight parameters exploiting the no-signaling condition. Indeed, note first that the three numbers needed to specify $P(a, b \mid x, y)$ can be chosen as being $P(a = 0 \mid x, y)$, $P(b = 0 \mid x, y)$ and $P(a = 0, b = 0 \mid x, y)$, since $P(a = 0, b = 1 \mid x, y) = P(a = 0 \mid x, y) - P(a = 0, b = 0 \mid x, y)$ because $P(a = 0 \mid x, y) = P(a = 0, b = 0 \mid x, y) + P(a = 0, b = 1 \mid x, y)$. Furthermore, $P(a = 1, b = 0 \mid x, y) = P(b = 0 \mid x, y) - P(a = 0, b = 0 \mid x, y)$ and $P(a = 1, b = 1 \mid x, y)$ follows by normalization. Because of no-signaling, $P(a = 0 \mid x, y) = P(a = 0 \mid x)$ for all x and $P(b = 0 \mid x, y) = P(b = 0 \mid y)$ for all y, that is Alice's marginal distribution does not depend on Bob's choice of input and viceversa.

Therefore, there are eight probabilities left that can be conveniently arranged as a table:

$$P \quad = \quad \begin{array}{c|cc} A \setminus B & P(b|0) & P(b|1) \\ \hline P(a|0) & P(a,b|0,0) & P(a,b|0,1) \\ P(a|1) & P(a,b|1,0) & P(a,b|1,1) \end{array} \tag{4.14}$$

Employing the restrictions between probabilities, this is equivalent to:

$$P \quad = \quad \begin{array}{c|cc} A \setminus B & P(b=0|0) & P(b=0|1) \\ \hline P(a=0|0) & P(0,0|0,0) & P(0,0|0,1) \\ P(a=0|0) & P(0,0|1,0) & P(0,0|1,1) \end{array} . \tag{4.15}$$

For instance, the probability distribution of a PR-box and the one associated to the best measurements on a maximally entangled state (see 2.3) read respectively

$$P_{PR} = \begin{array}{c|cc} & 1/2 & 1/2 \\ \hline 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{array} \quad , \quad P_{ME} = \begin{array}{c|cc} & 1/2 & 1/2 \\ \hline 1/2 & \frac{1+1/\sqrt{2}}{4} & \frac{1+1/\sqrt{2}}{4} \\ 1/2 & \frac{1+1/\sqrt{2}}{4} & \frac{1-1/\sqrt{2}}{4} \end{array} . \tag{4.16}$$

Since we have already derived the CHSH inequality in a different way, we can study here how the inequality looks like in this notation. Noting that $C_{xy} = 1 - 2[P(0,1|x,y) + P(1,0|x,y)] = 4P(0,0|x,y) - 2P(a=0|x) - 2P(b=0|y) + 1$, since $P(a=0|x) = P(0,0|x,y) + P(0,1|x,y)$ we find

$$\langle S\rangle = 4\left[P(0,0|0,0) + P(0,0|0,1) + P(0,0|1,0) - P(0,0|1,1) - P(a=0|0) - P(b=0|0)\right] + 2\,.$$

Remembering that the inequality is $-2 \leq \langle S\rangle \leq 2$, by simply rearranging the terms we find

$$-1 \leq \quad P(0,0|0,0) + P(0,0|0,1) + P(0,0|1,0) - P(0,0|1,1) - P(a=0|0) - P(b=0|0) \quad \leq 0$$

i.e. the inequality known as Clauser-Horne (CH) — which is strictly equivalent to CHSH, under the assumption of no-signaling. Written as a table, we have

$$T_{CH} = \begin{array}{c|cc} & -1 & 0 \\ \hline -1 & 1 & 1 \\ 0 & 1 & -1 \end{array} \tag{4.17}$$

and the inequality reads

$$-1 \leq \quad \langle T_{CH}, P\rangle \quad \leq 0 \tag{4.18}$$

where $\langle , \rangle$ represent term-by-term multiplication (as a scalar product). For instance, $\langle T_{CH}, P_{PR}\rangle = \frac{1}{2}$ and $\langle T_{CH}, P_{ME}\rangle = \frac{1}{\sqrt{2}} - \frac{1}{2}$.

Local deterministic points (vertices of the local polytope)

The next step consists in identifying all the local deterministic strategies. There are only four deterministic functions $a = f(x)$ from one bit to one bit: $f_1(x) = 0$, $f_2(x) = 1$, $f_3(x) = x$ and $f_4(x) = x \oplus 1$. Therefore, there are sixteen local deterministic strategies $L^{ij}_{xy}(a,b) = \delta_{a=f_i(x)}\delta_{b=f_j(y)}$. Let us write down explicitly:

$$L_{11} = \begin{array}{c|cc} & 1 & 1 \\ \hline 1 & 1 & 1 \\ 1 & 1 & 1 \end{array}, \; L_{12} = \begin{array}{c|cc} & 0 & 0 \\ \hline 1 & 0 & 0 \\ 1 & 0 & 0 \end{array}, \; L_{13} = \begin{array}{c|cc} & 1 & 0 \\ \hline 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}, \; L_{14} = \begin{array}{c|cc} & 0 & 1 \\ \hline 1 & 0 & 1 \\ 1 & 0 & 1 \end{array},$$

$$L_{21} = \begin{array}{c|cc} & 1 & 1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \; L_{22} = \begin{array}{c|cc} & 0 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \; L_{23} = \begin{array}{c|cc} & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \; L_{24} = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array},$$

$$L_{31} = \begin{array}{c|cc} & 1 & 1 \\ \hline 1 & 1 & 1 \\ 0 & 0 & 0 \end{array}, \; L_{32} = \begin{array}{c|cc} & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \; L_{33} = \begin{array}{c|cc} & 1 & 0 \\ \hline 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}, \; L_{34} = \begin{array}{c|cc} & 0 & 1 \\ \hline 1 & 0 & 1 \\ 0 & 0 & 0 \end{array},$$

$$L_{41} = \begin{array}{c|cc} & 1 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 1 & 1 \end{array}, \; L_{42} = \begin{array}{c|cc} & 0 & 0 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array}, \; L_{43} = \begin{array}{c|cc} & 1 & 0 \\ \hline 0 & 0 & 0 \\ 1 & 1 & 0 \end{array}, \; L_{44} = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array},$$

According to the theorem shown in Theorem 4.1, we know that the set of local distributions (distributions that can be obtained with pre-established strategies) is the convex set whose extremal points are the deterministic strategies, that forms a polytope, the local polytope. The extremal points of a polytope are called vertices.

Facets of the local polytope: Bell's inequalities

The vertices of the local polytope define its facets, i.e. the planes that bound the set. If a point is below the facet, the corresponding probability distribution can be reproduced with local variables; if a point is above the facet, it cannot. Therefore, the facets of the polytope are the geometric representation of Bell's inequalities. Actually, in addition to Bell's inequalities, there are many

facets that are trivial and can never be violated: conditions like $P(a,b|x,y) = 0$ or $P(a,b|x,y) = 1$ obviously define boundaries within which every local distribution must be found.

The local polytope for this case is a polygone in an 8-dimensional space, so its facets are 7-dimensional planes. One has to identify the sets of eight points that define one of these planes, then write the equation that define each plane. This can be done by brute force, but we just use some intuition and then quote the known result. The CH inequality indeed defines facets: we have $\langle T_{CH}, L\rangle = 0$ for $L \in \{L_{11}, L_{13}, L_{22}, L_{24}, L_{31}, L_{34}, L_{42}, L_{43}\}$; these eight points indeed define a plane of dimension 7. Above this facet, the most non-local point is the PR-box (4.7).

These local strategies can also be expressed in the following way[4]:

$$\begin{aligned} L_1^r &= \{a(x) = r, b(y) = r\} \\ L_2^r &= \{a(x) = x + r, b(y) = r\} \\ L_3^r &= \{a(x) = r, b(y) = y + r\} \\ L_4^r &= \{a(x) = x + r, b(y) = y + r + 1\} \end{aligned} \tag{4.19}$$

with the equivalences with the previous notation:

$$\begin{aligned} L_1^0 &= L_{11} \quad L_1^1 = L_{22} \\ L_2^0 &= L_{31} \quad L_2^1 = L_{42} \\ L_3^0 &= L_{13} \quad L_3^1 = L_{24} \\ L_4^0 &= L_{34} \quad L_4^1 = L_{43} \end{aligned} \tag{4.20}$$

Also, $T_{CH} \cdot L = -1$ for the other eight deterministic points: this facet is "opposite" to the previous one, with the local polytope between the two. The most non-local point above this facet is also a PR-box, the one defined by the rule $a \oplus b = xy \oplus 1$. Note that this PR-box is obtained from the "original" one by trivial local processing: e.g. Alice flips her outcome.

A simple argument of symmetry gives us immediately six other equivalent facets. Indeed, given a Bell inequality, a relabeling of the inputs and/or the outputs provides another Bell inequality. Their number is most easily counted by studying how many different rules for PR-boxes one can find, and these are obviously $a \oplus b = (x \oplus 1)y$, $a \oplus b = x(y \oplus 1)$ and $a \oplus b = (x \oplus 1)(y \oplus 1)$, with the corresponding opposite facets obtained as before by adding 1 on either side.

Now, it can be proved that there are no other Bell's inequalities for this case: only eight versions of CHSH, each with eight extremal points on the corresponding facet and one single PR-box on top. All these facets being equivalent to the others up to trivial relabeling of the inputs and/or the outputs, are CHSH inequalities.

The no-signaling polytope and the quantum set

We have studied at length the set of local distributions. Now we have to say a few words about the two other meaningful sets, namely the no-signaling distributions and the distributions that can be obtained with quantum physics. The image of the probability space is usually drawn[5] as in Fig. 4.3.

The no-signaling conditions define a convex set (if two distributions are no-signaling, any convex combination will also be no-signaling). It turns out that this set is also a polytope, i.e., it has a finite number of extremal points; obviously, it is called no-signaling polytope. All the local deterministic points are also extremal for the no-signaling polytope; but in addition to those, of course, there are some non-local ones, that can in principle be found[6]. For the simple example we studied, the only additional extremal points are the eight PR-boxes defined above.

The quantum set is also convex, but is not a polytope: it has an uncountable number of extremal points. Interestingly, at the moment of writing, the shape of this convex body has not been

[4] We introduce this notation because it's the one that we will employ in chapter 5 for describing the $d = 2$ protocol. The notation in arrays is more useful to generalize the protocol to the $d > 2$ case, thus its detailed exposition.

[5] This drawing is a poor representation of an 8-dimensional object.

[6] To characterize the facets of the no-signaling polytope, one has to consider that they are facets that satisfy the no-signaling condition.

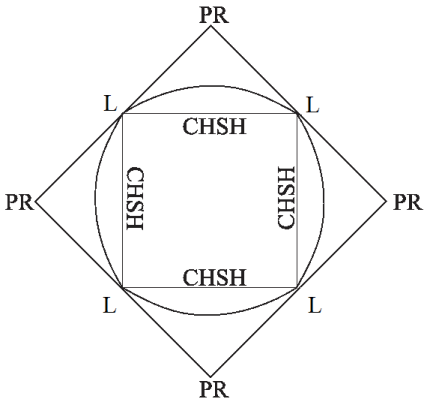


Figure 4.3: Representation of the local polytope, the set of quantum correlations and the no-signaling polytope, for the case under study (2 parties, 2 inputs and 2 outputs per party). The local polytope (inner square) is delimited by versions of the CHSH inequality; the quantum set (round body) exceeds the local polytope, and is contained in the no-signaling polytope (external square), whose extremal points are PR-boxes. Local deterministic points (L) are extremal points of all three sets (not obvious in the drawing, which is just a projection on a two-dimensional slice: the deterministic points do not lie on this slice).

characterized in full generality yet, not even for the simple case under study here. A necessary condition is the following for any probability distribution coming from quantum physics, the correlation coefficients must satisfy an inequality that reminds of CHSH, namely

$$|\arcsin(C_{xy}) + \arcsin(C_{xy'}) + \arcsin(C_{x'y}) - \arcsin(C_{x'y'})| \quad \leq \quad \pi\,. \tag{4.21}$$

However, this condition is provably not sufficient in general: there are probability distributions that satisfy this inequality but cannot be produced with quantum states [17].

4.2.4 Non-signaling polytope, in $d > 2$

We present in this section the generalization of the non-signaling polytope that will be employed for the explicit analysis of the protocol in the case $d = 3$ in chapter 5. This generalization of the previous results consists on binary inputs and d-nary outputs: $m_A = m_B = 2$, $n_A = n_B = d$; that is, $x, y \in \{0, 1\}$ and $a, b \in \{0, 1, ..., d-1\}$. Below, all the sums involving dits are to be computed modulo d.

While the full probability space is $4d^2$-dimensional, one can verify that only $4d(d-1)$ parameters are needed to characterize completely a no-signalling probability distribution — in other words, $D = 4d(d-1)$ is the dimension of the space in which the no-signalling and the local polytopes are embedded. We choose the $\{P(a|x),\ a = 0, 1, ..., d-2;\ x = 0, 1\}$ ($d-1$ numbers for each value of x), the $\{P(b|y), b = 0, 1, ..., d-2;\ y = 0, 1\}$ ($d-1$ numbers for each value of y), and the $\{P(a, b|x, y),\ a, b = 0, 1, ..., d-2;\ x, y = 0, 1\}$ ($(d-1)^2$ numbers for each value of x, y). This we arrange in arrays as follows, as in the $d = 2$ example:

$$P \quad = \quad \begin{array}{|c|c|c|} \hline A \setminus B & P(b|0) & P(b|1) \\ \hline P(a|0) & P(a, b|0, 0) & P(a, b|0, 1) \\ \hline P(a|1) & P(a, b|1, 0) & P(a, b|1, 1) \\ \hline \end{array} \tag{4.22}$$

Note that this array has $2(d-1)$ lines and as many columns: information on the values $a, b = d-1$ is redundant for all inputs because of no-signalling. Of course, there is no problem in working with the "full" array with $2d \times 2d$ if one finds it more convenient, provided the additional entries are filled consistently because these parameters are not free.

As in the case of $d = 2$, this notation will also be used for inequalities: in this case, the numbers in the arrays are the *coefficients* which multiply each probability in the expression of the inequality.

In order to describe this generalization of the non-signaling polytope, a more general type of Bell inequalities must be described.

CGLMP inequalities

We introduce a general Bell inequality for any possible number of outcomes d. This generalization, the so-called Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality, was first introduced in [21] and its explanation is crucial for the presentation of the protocol for $d = 3$ in the next chapter.
The idea behind the derivation of this inequalities is exploiting a logical constraint that all local variables must satisfy. Let's describe in the following lines the procedure derived in the original article. Alice's two possible measurements are labeled a_0, a_1 anb Bob's b_0, b_1. Each measurement has d possible outcomes $a_0, a_1, b_0, b_1 = 0, \ldots, d-1$. The local correlation can be described by d^4 probabilities c_{jklm} $(j, k, l, m = 0, \ldots, d-1)$ such that Alice's measurements a_0 and a_1 give outcomes j, k respectively, and equivalently for Bob. We have that $c_{jklm} \geq 0$ and $\sum_{jklm} c_{jklm} = 1$. The joint probabilities are of the form $P(a_0 = j, b_0 = l) = P(j, l|0, 0) = \sum_{km} c_{jklm}$ and similarly for the rest of measurements.
Given a choice of local variables $jklm$, the following differences are defined

$$\begin{aligned} r' &\equiv \quad b_0 - a_0 = l - j \\ s' &\equiv \quad a_1 - b_0 = k - l \\ t' &\equiv \quad b_1 - a_1 = m - k \\ u' &\equiv \quad a_0 - b_1 = j - m \end{aligned} \tag{4.23}$$

That gives the following constraint

$$r' + s' + t' + u' = 0 \tag{4.24}$$

The relation between three pairs of operators can be freely chosen but the fourth is constrained. The Bell inequalities here presented exploit the violation of such constraint. The simplest Bell expression derived in this way is:

$$\mathcal{B} \equiv P(a_0 = b_0) + P(b_0 = a_1) + P(a_1 = b_0) + P(b_1 = a_1 + 1) = \sum_{x,y} P(a + b = xy|x, y) \tag{4.25}$$

where we represent the probability $P(a_x = b_y + k)$ that the measurements a_x and b_y have outcomes that differ $k \mod d$:

$$P(a_x = b_y + k) \equiv P(a = b + k|x, y) \equiv \sum_{j=0}^{d-1} P(a = j, b = j + k \mod d|x, y) \tag{4.26}$$

Evaluating the differences between outcomes modulo d involves a symmetrization that is key to reducing the Bell inequalities to the logical constraint 4.24. Therefore, imposing this constraint implies that any choice of local variables only can satisfy three of the four relations appearing in 4.25 and consequently $\mathcal{B}(local) \leq 3$. However, nonlocal correlations can attain $\mathcal{B} = 4$ as they can satisfy all 4 relations. In the case $d = 2$, the inequality $\mathcal{B}(local) \leq 3$ is equivalent to the CHSH inequality described profusely in this memoir. This reformulation is useful for two reasons:

- In the following chapter we will employ this formulation of the CHSH inequality for $d = 2$ in order to measure the violation of the inequality in terms of a parameter that is crucial in the protocol described.
  In this case, the maximal violation is 4 and the violation that bounds the quantum region is $\mathcal{B}(quantum) \leq 2 + \sqrt{(2)}$
- It can generalize the CHSH equation to higher dimensions and represents already known Bell inequalities in this form and specify new Bell inequalities.

Without giving further details about its derivation, we give the following generalizations of these so-called CGLMP inequalities

$$\begin{aligned} \tilde{I}_3 = \quad & [P(a = b|0, 0) + P(b = a + 1|0, 1) + P(a = b|1, 1) + P(a = b|1, 0)] \\ & - [P(a = b - 1|0, 0) + P(a = b|0, 1) + P(a = b - 1|1, 1) + P(b = a - 1|1, 0)] \end{aligned} \tag{4.27}$$

The maximum value attainable for non-local theories is 4, since a non local theory could satisfy all the 4 relations with a + sign in the previous expression. For local correlations, the maximal violation is $\tilde{I}_3 \leq 2$. This is different to the local limit established in (4.25), due to the negative signs in (4.27) (only 3 relationships in (4.25) could be satisfied by local correlations, but in (4.27) if 3 with + sign are satisfied, one with - sign is also satisfied).
For $d = 2$, the inequality $\tilde{I}_3(local) \leq 2$ is equivalent to the CHSH inequality. However, the generalization of $\tilde{I}_3$ for higher dimensional outcomes, $\tilde{I}_d$, is more robust as it is expressed in [21]. The expression for this generalization is

$$\begin{aligned}\tilde{I}_d \;=\; & \sum_{k=0}^{[d/2]-1}\left(1-\frac{2k}{d-1}\right)\Big\{\big[P(a=b-k|0,0)+P(a=b+k|0,1)+P(a=b+k|1,0)+P(a=b-k-1|1,1)\big] \\ & -\big[P(a=b+k+1|0,0)+P(a=b-k-1|0,1)+P(a=b-k-1|1,0)+P(a=b+k|1,1)\big]\Big\} \qquad (4.28)\end{aligned}$$

This is the CGLMP inequality.

Non-signaling polytope, case d

To describe the non-local distributions that lie above the facets given by CGLMP inequalities and to explicit the structures that will be employed to develop the protocol in the next chapter, we consider a modified version of the CGLMP inequality (4.28), employing the notations already discussed in the case $d = 2$.

$$I_d \;=\; \begin{array}{|c|cccc|cccc|}\hline A\setminus B & -1 & -1 & \dots & -1 & 0 & 0 & \dots & 0 \\ \hline -1 & 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 & 1 & 1 & \dots & 1 \\ \hline 0 & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 & -1 & -1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 1 & 1 & \dots & 1 & -1 & -1 & \dots & -1 \\ \hline\end{array} \leq 0\,. \qquad (4.29)$$

The link between $\tilde{I}_d$ and this definition of I_d is provided by

$$I_d \;=\; \frac{d-1}{2d}\left(-2+\tilde{I}_d\right). \qquad (4.30)$$

This is a generalization of the transformation of CHSH inequality into CH seen before.

The highest violation of CGLMP is provided by the extremal point

$$PR_{2,d} \;=\; \frac{1}{d}\,\delta(b-a=xy)\,, \qquad (4.31)$$

whose corresponding array is

$$PR_{2,d} \;=\; \frac{1}{d}\;\begin{array}{|c|cccc|cccc|}\hline A\setminus B & 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \\ \hline 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \\ \hline 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots & \ddots & \ddots & 1 \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ \hline\end{array}\,. \qquad (4.32)$$

The violation of the inequality can be rapidly calculated by a term-by-term multiplication (a formal scalar product as describe in the case $d = 2$) of the two arrays (4.29) and (4.32), yielding

$$\langle I_d, PR_{2,d} \rangle \quad = \quad \frac{d-1}{d} \,. \tag{4.33}$$

However, $PR_{2,d}$ is not the only non-local extremal point which lies above a CGLMP facet: in fact, for all $d' < d$, there is at least one $PR_{2,d'}$ above the facet. For instance, a possible version of $PR_{2,2} \equiv PR$ reads (boldface $\mathbf{0}$ standing for matrices filled with zeros)

$$PR \quad = \quad \frac{1}{2} \left[\begin{array}{c|cc|cc} A \setminus B & 1 \;\; 1 & \mathbf{0} & 1 \;\; 1 & \mathbf{0} \\ \hline 1 & 1 \;\; 0 & & 1 \;\; 0 & \\ 1 & 0 \;\; 1 & \mathbf{0} & 0 \;\; 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \;\; \mathbf{0} & \mathbf{0} & \mathbf{0} \;\; \mathbf{0} & \mathbf{0} \\ \hline 1 & 1 \;\; 0 & & 0 \;\; 1 & \\ 1 & 0 \;\; 1 & \mathbf{0} & 1 \;\; 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \;\; \mathbf{0} & \mathbf{0} & \mathbf{0} \;\; \mathbf{0} & \mathbf{0} \end{array}\right] , \tag{4.34}$$

whence a violation $\langle I_d, PR \rangle = \frac{1}{2}$.

For $d = 3$, we shall give below (5.4.2) some additional elements on the structure of the no-signalling polytope.

A slice in the non-local region

As we said, a no-signalling probability distribution is characterized by $4d(d-1)$ parameters. However, when one reviews the results obtained for the CGLMP inequality in the context of quantum physics [7], one finds that the probability distributions associated to the optimal settings belong to a very symmetric family. Specifically, these distributions are such that

1. for fixed inputs x and y, $P(a, b|x, y)$ depends only on $\Delta = a - b$; and
2. the probabilities for the different inputs are related as $P(\Delta|0, 0) = P(-\Delta|0, 1) = P(-\Delta|1, 0) = P(\Delta - 1|1, 1)$.

Compactly:

$$P(a, b = a - \Delta|x, y) \quad = \quad \frac{1}{d}\, p_f \tag{4.35}$$

with $f = (-1)^{x+y}\Delta + xy$ and $p_f = \sum_a P(a, b = a - f|0, 0)$. The corresponding array is

$$P \quad = \quad \frac{1}{d} \left[\begin{array}{c|cccc|cccc} A \setminus B & 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \\ \hline 1 & p_0 & p_{-1} & \dots & p_2 & p_0 & p_1 & \dots & p_{-2} \\ 1 & p_1 & p_0 & \dots & p_3 & p_{-1} & p_0 & \dots & p_{-3} \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 1 & p_{-2} & p_{-3} & \dots & p_0 & p_2 & p_3 & \dots & p_0 \\ \hline 1 & p_0 & p_1 & \dots & p_{-2} & p_1 & p_0 & \dots & p_3 \\ 1 & p_{-1} & p_0 & \dots & p_{-3} & p_2 & p_1 & \dots & p_4 \\ \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 1 & p_2 & p_3 & \dots & p_0 & p_{-1} & p_{-2} & \dots & p_1 \end{array}\right] . \tag{4.36}$$

This family defines a *slice* in the no-signalling polytope. Note that all the marginals are equal, that is, $P(a|x) = P(b|y) = \frac{1}{d}$. Moreover, the d numbers p_f define uniquely and completely a point P in the slice; thus, given the constraint $\sum_f p_f = 1$, the slice defined by (5.53) is $(d-1)$-dimensional. A single extremal non-local point belongs to the slice, namely $PR_{2,d}$, obtained by setting $p_0 = 1$ (4.31); in fact, none of the $PR_{d'}$ with $d' < d$ has the correct marginals.

[7] Details are provided in Appendix B of [18] and are out of the scope of this memoir

In the following chapter, a depolarization procedure that maps any probability distribution onto this slice will be described. This depolarization protocol consists on local operation and public communication and guarantees that the violation of the correspondent CGLMP inequality is conserved (that is $\langle I_d, P\rangle = \langle I_d, P_2\rangle$, where P_2 is the mapping into the slice).

The derivation of this slice is fundamental to the protocol developed in the next chapter for $d > 2$ outcomes.

4.3 Non locality as an information resource

We expose the formalism to study non-signaling theories and describe the results that allows us to derive information theory results equivalent to the developed for quantum information theory.
This section is based in the paper [15]. The notation employed is the same of the rest of the chapter and some results will be detailed as they are important foundations for the next chapter. Along the following sections, the Bell inequalities for correlations will be supposed to be normalized, that is that a Bell inequality $\tilde{\mathcal{B}}$ is transformed linearly as $\mathcal{B} = \alpha\tilde{\mathcal{B}} + \beta$ in order to have the local correlations bounded by 0, $\mathcal{B}(P_{local}) \leq 0$ and the maximal violation compatible with no signaling expressed as $\mathcal{B}(P_{max}) = 1$ [8]

4.3.1 Monogamy, m-shareability and No-cloning

Classical correlations can be shared by an indefinite number of parties but quantum correlation cannot. This is called monogamy of entanglement and can be generalized to all non-signaling theories. The first result to consider is that all Bell inequalities with a unique distribution being the maximal violation consistent with no signaling have monogamy constraints. That is, if $\mathcal{B}$ is a Bell inequality with a unique maximal violator $P_{\max}$ and Alice and Bob maximally violate a Bell inequality, $\mathcal{B}(P(a,b|x,y)) = 1$, then Alice and, say, Claire are completely uncorrelated. To derive this result, one must take into account that $P_{\max}$ must be an extremal point of the Alice-Bob polytope, since all Bell inequalities are linear in P, that is $\mathcal{B}(\alpha P_1 + \beta P_2) = \alpha\mathcal{B}(P_1) + \beta\mathcal{B}(P_2)$. If this were not the case, then the maximal violator would not be unique. Decomposing $P_{\max}$ in the following way, applying definition of conditional probabilities and no signaling principle:

$$P_{\max}(a,b|x,y) \quad = \sum_c P(a,b,c|x,y,z) = \sum_c P(a,b|x,y,z,c)P(c|z)\,\forall z \tag{4.37}$$

But being $P_{\max}(a,b|x,y)$ extremal, the decomposition is composed of only one term and thus Claire is uncorrelated with Alice and Bob.
A bipartite probability distribution $P(a,b|x,y)$ is said to be m-shareable with respect to Bob if there exists an $(m+1)$-partite distribution, $P(a,b_1,\ldots,b_m|x,y_1,\ldots,y_m)$ symmetric respect to $(b_i,y_i)\,\forall i \in \{1,\ldots,m\}$, whose marginals $P(a,b_i|x,y_i)$ are equal to the original distribution $P(a,b|x,y)$.

The following result establishes the link between shareability and locality

Proposition 4.1 *If $P(a,b|x,y)$ is m-shareable with respect to Bob, then it satisfies all Bell inequalities with m different values of the input y*

Proof: The proof is based in the construction of a local model for $P(a,b|x,y)$ having y constrained to $y = 1,\ldots,m$. The distribution $P(a,b_1,\ldots,b_m|x,y_1,\ldots,y_m)$ exists and the probabilities $P(b_1,\ldots,b_m|y_1,\ldots,y_m)$ and the conditional $P(a|x,b_1,\ldots,b_m,y_1,\ldots,y_m)$ do also exist. The local variable defined as $\lambda = (b_1,\ldots,b_m)$ is the information shared by the parties, when the corresponding inputs are fixed as $y_1 = 1,\ldots,y_m = m$. Employing the decomposition

$$P(a,b|x,y) = \sum_{b_1,\ldots,b_m} P(b_1,\ldots,b_m|y_1 = 1,\ldots,y_m = m) \tag{4.38}$$

$$P(a|x,b_1,\ldots,b_m,y_1 = 1,\ldots,y_m = m)\delta_{b,b_y} \tag{4.39}$$

where δ_{b,b_y} is 1 for the variables b and y appearing in $P(a,b|x,y)$. This decomposition is the definition of local correlations $P(a,b|x,y) = \sum_\lambda P(\lambda)P(a|x,\lambda)P(b|y,\lambda)$

[8] This is exactly the procedure employed to derive the CGLMP inequalities I_d.

□

The consequence of the equivalence between shareability and locality is that any non signaling theory has a no cloning theorem as in the Quantum Mechanics case. The no cloning theorem of Quantum Mechanics, which is one of the cornerstones of Quantum Information Theory is explained as a consequence of nonorthogonality of quantum states and linearity of evolution (this is the way it was shown in chapter 2). The existence of a no cloning theorem for more general no signaling theories is based on the no shareability of non local correlations.

Proposition 4.2 *All non-signaling theories predicting the violation of Bell inequalities have a no-cloning theorem*

Proof: Suppose there exist a machine to which we can input a physical system in a given state and it outputs two systems in exactly the same state as the original one. This is a perfect cloning machine. If Alice and Bob share a nonlocal distribution $P(a, b|x, y)$, let's assume the following situation and arrive to a contradiction.

- Alice chooses input x_0 and obtains output a_0
- Bob creates m clones of its original system

From the point of view of an observer that first witnesses Alice's measure, the description of Bob's system is $P(b|y, x_0, a_0)$. After cloning, this distribution becomes

$$P(b|y, x_0, a_0) \to P(b_1, \ldots, b_m|y_1 \ldots, y_m, x_0, a_0) \tag{4.40}$$

The marginal $P(b_i|y_i, x_0, a_0)$ is the same as the original $P(b|y, x_0, a_0)$. For an observer that first sees Bob's operation, the distribution is

$$P(a, b_1, \ldots, b_m|x, y_1 \ldots, y_m) \tag{4.41}$$

But both descriptions have to give consistent predictions and thus must be the same up to conditioning on a. This implies that the original distribution is shareable and thus local. □

The existence of both monogamy and impossibility of cloning is related to the concept of secrecy extraction. In the next section we explore this relationship.

4.3.2 Nonlocality and secrecy extraction

The relationship between general non-locality and secrecy extends the results that are known between entanglement and secrecy (as the EPR protocol described in Chapter 2). That general non signaling theories exhibit the same properties as quantum mechanics, gives the intuition that they can be employed as a source of secrecy between two parties. If two honest parties share correlations with certain degree of monogamy, they can estimate and bound their correlations with an adversary, the eavesdropper. The correlations between the two parties and the adversary are described by a probability distribution P_{ABE} where E stands for the eavesdropper random variable, with the notation employed in chapter 3.

This distribution contains secrecy if it cannot be generated by local operations and public communication (LOPC), that is, it requires private channels or secret bits for its formation. We say that P_{ABE} contains secrecy if there doesn't exist a stochastic map $E \to E'$ such that

$$P_{AB|E'} = P_{A|E'} P_{B|E'}$$

This does not automatically imply that the distribution between the three parties can later be employed to generate a secret key, as there are probability distributions containing secrecy from which a secret key cannot be distilled (the situation where there is bound information [35]). Alice and Bob share a distribution $P(a, b|x, y)$ and decide their inputs according to uniform distributions $P(x) = \frac{1}{X}$ and $P(y) = \frac{1}{Y}$. The random variables $A = (a, x)$ and $B = (b, y)$ are correlated as:

$$P_{AB} = P(a, b|x, y)\frac{1}{XY} \tag{4.42}$$

These correlations can be extended to the distribution including the eavesdropper Eve assuming non-signaling into $P(a,b,e|x,y)$ and therefore

$$P_{ABE} = P(a,b,e|x,y)\frac{1}{XY} \tag{4.43}$$

where $E = e$. We say that $P(a,b|x,y)$ contains secrecy if all its associated P_{ABE} contain secrecy. The equivalence between no secrecy content and locality of correlations is derived from the fact that a probability distribution $P(a,b|x,y)$ is local when there is an extension $P(a,b,e|x,y)$ such that

$$P(a,b|x,y,e) = P(a|x,e)P(b|y,e) \tag{4.44}$$

As already stated the bipartite distribution doesn't have secrecy content if there is an extension

$$P_{AB|E} = P_{A|E}P_{B|E}$$

Both equations are equivalent as the arbitrariness of the extension $P(a,b,e|x,y)$ includes any transformation $E \to E'$. What we are implying here is that P_{ABE} has no secrecy if and only if there exists an extension of $P(a,b|x,y)$ satisfying 4.44, that is that the distribution is local.

We have described in this chapter a generalization of non local correlations and Bell inequalities and the formalism of the non-signaling polytope. The results presented are the foundations in which the protocol described in the next chapter is based.

The next chapter deals with the exposition of a protocol based on non signaling correlations in which a secret key can be extracted.

Chapter 5

Secrecy extraction from non-signaling correlations

In the previous chapter, the non-signaling polytope was introduced as an information resource. In the present chapter we develop a key distribution protocol based on the violation of a Bell inequality and study its properties.

5.1 Introduction

In this chapter we present a cryptographic protocol initially developed by A. Acín, N. Gisin and LL. Masanes ([2]) which is based on the violation of a CHSH-type inequality. It has been my job during the development of this project to study interesting characteristics of this protocol -such as positivity of intrinsic information, minimum bound of key rate in individual attacks in its $d = 2$ and $d = 3$ versions and advantage distillation techniques.

The idea of a protocol which is based on physical principles beyond quantum formalism, is motivated by the fact that its assumptions are less restrictives. QKD is not only based on the validity of quantum formalism, there is also an intrinsic extra assumption: Alice and Bob know how they have done their measures -i.e. how they obtained their correlations $P(a, b \mid x, y)$. As in the photon-polarization implementation of BB84 protocol provided in chapter 2, Alice knows she sends photons in two polarization bases, z and x, that are measured by Bob in the same bases.

As it was explained in the introduction, this assumption means that the dimension of the Hilbert space of the state employed in generating the key is under control and not a resource that could be exploited by the eavesdropper.

In order to circumvent this assumption and as the initial steps in developing Device Independent Quantum Key Distribution, a protocol assuming just the violation of a Bell inequality was proposed. The motivation behind this protocol was a non-signaling protocol proposed in [5]. However, this protocol was not suited to extract a secret key.

5.2 Secrecy of probability distributions and individual eavesdropping strategies

Alice and Bob have realized the measurement several times and share an arbitrary number of realizations that are distributed as $P(a, b|x, y)$. By revealing some of their results, they can establish the degree of violation of the CHSH inequality (thus if the probability distribution that they share is inside the local polytope or in the non-local region). The goal is to study whether Alice and Bob can extract secrecy out of their data with this knowledge only.

Let's consider the BB84 protocol described in chapter 2 under this point of view. In this protocol, $a, b \in \{0, 1\}$ and $x, y \in \{X, Z\}$ are both binary. In the absence of errors, perfect correlations when $x = y$ and no correlations when $x \neq y$ are distributed. Expressed in terms of conditioned

probabilities, that is:

$$\begin{aligned} P(0,0|X,X) = P(1,1|X,X) &= \frac{1}{2} \\ P(0,0|Z,Z) = P(1,1|Z,Z) &= \frac{1}{2} \\ P(a,b|X,Z) = p(a,b|Z,X) &= \frac{1}{4} \end{aligned} \tag{5.1}$$

If Alice and Bob have obtained their results by measuring qubits, the correlations obtained generate secrecy under the assumption that the eavesdropper is limited only by the laws of quantum physics. However, the distribution can also be obtained with shared randomness: if Alice and Bob would share randomly distributed pairs of classical bits (r_X, r_Z), they simply have to output r_Z (respectively r_X) if they are asked to measure Z (respectively X). The additional assumption on the physical realization is what makes the protocol secure, as Alice and Bob are both measuring a qubit. The correlations that they share are not secure per se. The exploration that follows is motivated to provide correlations that are secure by themselves without having to make any extra assumption in the way they were generated or shared.

This first step in studying a protocol based in the violation of Bell inequalities was made with the following assumptions:

- the eavesdropper Eve is not even limited by quantum physics, but only by the no-signalling constraint
- the eavesdropper Eve is limited to adopt an individual strategy (that is, she cannot correlate her outputs)

Let's detail each one of the assumptions.

The generation of secrecy from non-local correlations was easier to analyze by considering that the eavesdropper Eve is only limited by the non-signaling constraint. This "supra-quantum" property of the eavesdropper means that Eve can distribute any many-instances probability distribution $P(\vec{a},\vec{b}|\vec{x},\vec{y})$ that lies within the no-signalling polytope (including the region beyond the quantum limit); Alice and Bob have the freedom of choosing their sequence of measurements, $\vec{x}$ and $\vec{y}$ respectively, and will obtain the corresponding outcomes. This is in fact the scheme described in the Introduction. This assumption is indeed conservative (although it was easier to study it than the quantum case), if a secret key can be extracted in this case, with a more realistic adversary the secret key extraction will be at least as efficient.

As we already described in the previous chapter, the secrecy in this case relies in the non-locality of the correlations.

Barrett, Hardy and Kent [5] showed an example of a protocol, in which quantum correlations can provide secrecy against the most powerful attack by a no-signaling Eve. This is the first example that one can achieve security even against a supra-quantum Eve, showing that security in key distribution arises from general features of no-signalling distributions rather than from the specificities of the Hilbert space structure. However, their example had important limitations: actually, it provides a protocol to distribute a single secret bit (hence zero key rate) in the case when Alice and Bob share correlations that can be ascribed to noiseless quantum states.

The work here presented provides an alternative to the above mentioned protocol but, as already stated, limiting the eavesdropper to adopt an individual strategy. The classification of different kinds of attacks from the point of view of the eavesdropper is in synthesis as follows:

- Individual attacks
  - Eve sends independent identically distributed signals
  - She tries to guess each bit of the raw key without performing correlations between instances (after classical post-processing between the parties). At the beginning of the classical post processing, Alice, Bob and Eve share a probability distribution of classical symbols.

- Collective attacks
  - Eve sends independent identically distributed signals
  - She tries to guess the final key and can correlate different instances of measurements.
- General or coherent attacks
  - Eve sends the most general signals (she may entangle systems, modify her attack according to some measurement)
  - She tries to guess the final key.

We will center the following discussions in individual attacks. This means that Eve follows the same procedure for each instance of measurement, she is not allowed to correlate different instances. Moreover, Eve is asked to put her input z before any error correction and privacy amplification. Consequently, any individual attack is described of a three-partite probability distribution $P(a,b,e|x,y,z)$ such that

$$P(a,b|x,y) \quad = \quad \sum_e P(e|z)\, P(a,b|x,y,e,z)\,. \tag{5.2}$$

Note that Eve is also limited by no-signalling, that is why the left hand side does not depend on z. One can see that this is an individual attack by looking at it as follows: when Eve gets outcome e out of her input z, she sends out the point $P(a,b|x,y,e,z)$.

Now we demonstrate two similar, important results about individual eavesdropping strategies related to her strategies involving the non-signaling polytope. These results make easier to analyze the probability distributions involved in the protocol.

Theorem 5.1 *Eve can limit herself in sending out extremal points of the no-signalling polytope.*

Proof: Suppose that an attack is defined, in which one of the $P(a,b|x,y,e,z)$ is not an extremal point. Then, this point can be itself decomposed on extremal points:

$$P(a,b|x,y,e,z) = \sum_\lambda P(\lambda) P(a,b|x,y,e,z,\lambda)$$

where the $P(a,b|x,y,e,z,\lambda)$ are all extremal. But the knowledge of λ must be given to Eve: by redefining Eve's symbol as $(e,\lambda) \to e'$, we have an attack which is as powerful as the one we started from, and is of the form (5.2) while having only extreme points in the decomposition. □

Theorem 5.2 *Suppose that Alice and Bob can transform $P(a,b|x,y)$ into $\tilde{P}(a,b|x,y)$ by using only local operations and public communication independent of a,b,x,y. Then there exist a purification of $\tilde{P}(a,b|x,y)$ that gives Eve as much information as the best purification of $P(a,b|x,y)$.*

Proof: Suppose (5.2) is the best purification of P from Eve's point of view; for clarity, let's use Theorem 1 to say that the $P(a,b|x,y,e,z)$ are extremal points. The procedure of Alice and Bob can be described as follows: for each realization of the variables (a,b,x,y), Alice draws a random number and reveals publicly its value j; then, she and Bob apply the local transformation T_j on which they have previously agreed, transforming $x \to X_j$, $a \to A_j$ etc. Since there is no correlation between j and (e,z), each extremal point $P_\epsilon(a,b|x,y,e,z)$ is transformed into

$$\begin{aligned} \tilde{P}(a,b|x,y,e,z) \quad &= \quad \sum_j P(j) P(A_j,B_j|X_j,Y_j,e,z) \\ &= \quad \sum_{j,\epsilon} P(j) P(\epsilon|e,j) P(a,b|x,y,\epsilon,z)\,. \end{aligned} \tag{5.3}$$

Consequently, $\tilde{P}(a,b|x,y)$ is a mixture of the extremal points $P(a,b|x,y,\epsilon,z)$ with weight $P(\epsilon|z) = \sum_{j,e} P(e|z)P(j)P(\epsilon|e,j)$. To conclude the proof, just notice that Eve has been able to follow the full procedure, because she has learnt j and the list of the T_j is publicly known. Thus, there exist a decomposition of $\tilde{P}(a,b|x,y)$ onto extremal points that gives Eve as much information as the best decomposition of $P(a,b|x,y)$. □

5.3 CHSH protocol

We present here the version of the protocol for binary outcomes (that can be implemented with qubits), and explore the results and extract some characteristics. In the case of qubits, we will talk about CHSH protocol. We have already characterized the non-signaling polytope for this case in chapter 4. We take the following representative of the CHSH inequality

$$\mathcal{B} = \sum_{x,y=0}^{1} P(a+b=xy \mid xy) \leq 3 \tag{5.4}$$

based in the CGLMP inequality in (4.25).

In order to analyze the properties of the protocol, we are going to explicitly express the probability distribution in terms of each of the vertices of the no-signaling polytope. To study the possibility of secret key extraction, we can restrict ourselves to the non-local region that lies above the facet of the local polytope determined by the local strategies in 4.19 and the PR-Box (4.9) that is above the the facet given by CHSH inequalities. Any point in this sector, by definition, can be decomposed as a convex combination of the PR-box and some of the eight deterministic strategies on the facet

$$\begin{aligned} L_1^r &= \{a(x)=r, b(y)=r\} \\ L_2^r &= \{a(x)=x+r, b(y)=r\} \\ L_3^r &= \{a(x)=r, b(y)=y+r\} \\ L_4^r &= \{a(x)=x+r, b(y)=y+r+1\} \end{aligned} \tag{5.5}$$

described in the previous chapter. As shown in the previous theorems, without loss of generality, Eve distributes these nine strategies. The notation employed in this section is different than the developed when presenting the no-signaling polytope in the case $d=2$ in chapter 4, but it is more suited to the calculations that we will develop onwards.

1. We shall write p_{NL} (for "non-local") the probability that Eve sends the PR-box to Alice and Bob;

2. and p_j^r the probability that Eve sends the deterministic strategy L_j^r. We shall also write $p_L = \sum_{j,r} p_j^r = 1 - p_{NL}$.

The statistics generated by Eve sending the extremal points are summarized in the Table 5.1.

Let's give some examples about the information given by the table. for example $P(a=b=0|x=y=0) = p_{NL}/2 + p_1^0 + p_2^0 + p_3^0$. To obtain the $P(a,b,x,y)$, one must multiply the entries of the Table by $P(x)P(y)$. Since we are supposing that the extremal points are sent to Alice and Bob by Eve, the label of each point can be considered also as Eve's symbol.

Finally, we note that if we substitute the probabilities given by Table 5.1 in the CHSH inequality (5.4), one finds

$$p_{NL} \;=\; CHSH \,-\, 3\,. \tag{5.6}$$

This means that the value p_{NL} measures directly the violation of the CHSH inequality and therefore is indicating the degree of nonlocality. It follows that for the quantum region, the violation is bounded by

$$p_{NL} \;\leq\; \sqrt{2} \,-\, 1 \approx 0.414. \tag{5.7}$$

The probability distribution that reaches this maximum is the one obtained by measuring observables on a maximally entangled state, whose probability distribution is $P(a+b=xy|x,y) = \frac{1+\frac{1}{\sqrt{2}}}{2}$. Whenever Eve distributes the PR-box, she has no information at all about the bits received by Alice and Bob, because of the monogamy of those correlations. On the contrary, when she distributes a deterministic strategy, she has some information, depending on the cryptographic protocol.

$A \setminus B$	$y=0,b=0$	$y=0,b=1$	$y=1,b=0$	$y=1,b=1$
$x=0$, $a=0$	$p_{NL}/2$ (PR) p_1^0 (L_1^0) p_2^0 (L_2^0) p_3^0 (L_3^0)	p_4^0 (L_4^0)	$p_{NL}/2$ (PR) p_1^0 (L_1^0) p_2^0 (L_2^0) p_4^0 (L_4^0)	p_3^0 (L_3^0)
$x=0$, $a=1$	p_4^1 (L_4^1)	$p_{NL}/2$ (PR) p_1^1 (L_1^1) p_2^1 (L_2^1) p_3^1 (L_3^1)	p_3^1 (L_3^1)	$p_{NL}/2$ (PR) p_1^1 (L_1^1) p_2^1 (L_2^1) p_4^1 (L_4^1)
$x=1$, $a=0$	$p_{NL}/2$ (PR) p_1^0 (L_1^0) p_3^0 (L_3^0) p_4^1 (L_4^1)	p_2^1 (L_2^1)	p_1^0 (L_1^0)	$p_{NL}/2$ (PR) p_2^1 (L_2^1) p_3^0 (L_3^0) p_4^1 (L_4^1)
$x=1$, $a=1$	p_2^0 (L_2^0)	$p_{NL}/2$ (PR) p_1^1 (L_1^1) p_3^1 (L_3^1) p_4^0 (L_4^0)	$p_{NL}/2$ (PR) p_2^0 (L_2^0) p_3^1 (L_3^1) p_4^0 (L_4^0)	p_1^1 (L_1^1)

Table 5.1: Table of the distribution Alice-Bob-Eve for the raw data case $d = 2$. The entries are the $P(a, b|x, y)$. In parentheses, we indicate Eve's symbol from the no-signaling polytope.

In order to extract a secret key from the distribution presented in Table 5.1, a protocol was proposed which was a reasonable candidate for optimality in secret key generation. The premises of a good protocol are to maximize the correlations between Alice and Bob at the same time that the information received by Eve is reduced. Observing the raw data, it can be noticed that Alice and Bob are highly anti-correlated when $x = y = 1$: it is thus natural to devise a procedure that allows them to transform these anti-correlations in correlations, revealing the least information in the public channel.

The following steps are the so-denominated CHSH protocol analyzed in this memoir:

1. *Distribution.* Alice and Bob repeat the measurement procedure on arbitrarily many instances and collect their data.

2. *Parameter estimation.* By revealing publicly some of their results, they estimate the parameters of their distribution, in particular, they measure the fraction p_{NL} of intrinsically non-local correlations.

3. *Pseudo-Sifting.* For each instance, Alice reveals the measurement she has performed ($x = 0$ or $x = 1$). Whenever Alice declares $x = 1$ and Bob has chosen $y = 1$, Bob flips his bit. Bob does not reveal the measurement he has performed. This is the procedure which transforms anti-correlations into correlations while revealing the smallest amount of information on the public channel. We call it pseudo-sifting, because it enters in the protocol at the same place as sifting occurs in other protocols, but here all the items are kept.

4. *Classical processing.* Here, classical techniques of secrecy extraction are employed in order to generate a key. The details depend on whether one considers one-way post-processing ("error correction and privacy amplification", efficient in terms of secret key rate) or two-way post-processing ("advantage distillation", inefficient for small errors but tolerating larger errors). The two cases will be analyzed in detail in the following sections.

Once pseudo sifting is performed, denoting the probability of each of Bob's measurements as $\xi_j = P(y = j)$, the Alice-Bob-Eve probability distribution can be written as it is shown in Table 5.2. In this table, the values of the probabilities of each symbol are split for each value of x. In order to understand the content of the distributions, let's analyze an example. If Eve sends out the deterministic point L_1^0 and Alice announces in the pseudo-sifting phase $x = 0$ then Eve knows for sure both Alice's and Bob's outcomes, in this particular case $a = 0$ and $b = 0$. For $x = 0$ it is the only result given by the strategy L_1^0. However, in the case that Eve sends out L_3^0 and Alice announces $x = 0$, Eve still knows for sure Alice's outcome ($a = 0$), but Bob's symbol is totally

dependent on his input, that is $b = 0$ if $y = 0$, $b = 1$ if $y = 1$. For $x = 1$, the roles are reversed L_3^0 produces only $a = 0$ and $b = 0$; while L_1^0 gives $a = 0$, $b = 0$ if $y = 0$ and $b = 1$ if $y = 1$.

All local points present the same behavior:

(i) Alice's outcome a is always known to Eve, because the setting used by Alice is publicly known.

(ii) If a local point employed by Eve gives her full information about Bob's outcome b when $x = 0$, the same point leaves her uncertain about b when $x = 1$ and viceversa.

Eve's uncertainty is maximal when Alice's and Bob's settings are chosen at random, therefore we set from now on

$$P(x = i) = \frac{1}{2}\, P(y = j) \;\equiv\; \xi_j \quad = \quad \frac{1}{2}\,. \tag{5.8}$$

This feature that the change of setting of measurement enlarges Eve's uncertainty is different from what is found in a traditional quantum cryptography setup. In this latter case, Eve's information does not depend on the frequency with which each setting is used, and in fact Alice and Bob can use almost always the same setting, provided they use the other one(s) sometimes in order to check coherence.

The pseudo-sifting phase of the protocol corrects the anti-correlation in the case $x = y = 1$ and although it increases information about Alice's symbol in the case of deterministic strategies, it keeps some uncertainty for Eve on Bob's result. The partial information about the setting is crucial to this point, if Alice and Bob would reveal both their settings, x and y, Eve would have total information on their respective symbols for all the deterministic strategies L_j^r. Table 5.2 contains the probability distribution Alice-Bob-Eve after pseudo-sifting. It is necessary to study the extraction of a secret key from this distribution using classical pre- and post-processing methods.

$\mathbf{x = 0}$	$b = 0$	$b = 1$
$a = 0$	$p_{NL}/2\ (PR)$ $p_1^0\ (L_1^0)$ $p_2^0\ (L_2^0)$ $p_3^0\xi_0\ (L_3^0)$ $p_4^0\xi_1\ (L_4^0)$	 $p_3^0\xi_1\ (L_3^0)$ $p_4^0\xi_0\ (L_4^0)$
$a = 1$	 $p_3^1\xi_1\ (L_3^1)$ $p_4^1\xi_0\ (L_4^1)$	$p_{NL}/2\ (PR)$ $p_1^1\ (L_1^1)$ $p_2^1\ (L_2^1)$ $p_3^1\xi_0\ (L_3^1)$ $p_4^1\xi_1\ (L_4^1)$
$\mathbf{x = 1}$	$b = 0$	$b = 1$
$a = 0$	$p_{NL}/2\ (PR)$ $p_1^0\xi_0\ (L_1^0)$ $p_2^1\xi_1\ (L_2^1)$ $p_3^0\ (L_3^0)$ $p_4^1\ (L_4^1)$	 $p_1^0\xi_1\ (L_1^0)$ $p_2^1\xi_0\ (L_2^1)$
$a = 1$	 $p_1^1\xi_1\ (L_1^1)$ $p_2^0\xi_0\ (L_2^0)$	$p_{NL}/2\ (PR)$ $p_1^1\xi_0\ (L_1^1)$ $p_2^0\xi_1\ (L_2^0)$ $p_1^3\ (L_1^3)$ $p_4^0\ (L_4^0)$

Table 5.2: Probability distributions Alice-Bob-Eve for the data sifted according to the CHSH protocol, conditioned to the knowledge of $x = 0$ or $x = 1$.

5.3.1 Uncertainty relations in the probability distribution

One feature that is worth noting about the probability distribution that is created by the protocol is that the eavesdropper gains information on a "basis" at the expense of increasing its errors in the

complementary one. This feature is similar to what happens in a traditional quantum cryptography setting (the choice of a basis increases Eve's information if this setting is chosen, as we saw in the explanation of the BB84 protocol in chapter 2).

Assuming maximal uncertainty in the election of settings, $\xi_0 = \xi_1 = \frac{1}{2}$, The probabilities $p(a \neq b|x) = e_{AB|x}$ of error between Alice and Bob when $x = 0$ and $x = 1$ are respectively

$$e_{AB|0} = \frac{1}{2}\left(p_3^0 + p_3^1 + p_4^0 + p_4^1\right), \tag{5.9}$$

$$e_{AB|1} = \frac{1}{2}\left(p_1^0 + p_1^1 + p_2^0 + p_2^1\right). \tag{5.10}$$

Eve's uncertainty on Bob's symbol, measured by conditional Shannon entropy, is obtained employing the formula

$$H(B|E, x = 0) = \sum_{b,e} P(b, e, x = 0) \log \frac{P(e, x = 0)}{P(b, e, x = 0)}$$

where e represents each of the strategies distributed.

$$H(B|E, x = 0) = 1 - \left(p_1^0 + p_1^1 + p_2^0 + p_2^1\right) \tag{5.11}$$

$$H(B|E, x = 1) = 1 - \left(p_3^0 + p_3^1 + p_4^0 + p_4^1\right). \tag{5.12}$$

Thus, the following uncertainty relation appear as a result of the protocol

$$H(B|E, x) = 1 - 2e_{AB|x+1}. \tag{5.13}$$

In terms of mutual information it can be expressed as

$$I(B : E|x = 0) = h\left(\frac{1}{2} + \frac{p_1^1 - p_1^0 + p_2^1 - p_2^0}{2}\right) - 1 + p_1^0 + p_1^1 + p_2^0 + p_2^1 \tag{5.14}$$

$$I(B : E|x = 1) = h\left(\frac{1}{2} + \frac{p_3^1 - p_3^0 + p_4^1 - p_4^0}{2}\right) - 1 + p_3^0 + p_3^1 + p_4^0 + p_4^1. \tag{5.15}$$

We can write this mutual information shortly as:

$$I(B : E|x) = h(\frac{1}{2} + \frac{p_{2x+1}^1 - p_{2x+1}^0 + p_{2x+2}^1 - p_{2x+2}^0}{2}) - 1 + 2e_{AB|x+1}. \tag{5.16}$$

This relationship between the information retrieved by Eve and the local strategies determined is explained analyzing Table 5.2. The pseudo-sifting phase aims to extract optimal correlation from the non-local strategy (distributing the PR-box, noted as PR) by reversing the anti-correlation of $x = y = 1$. But the consequence in the local deterministic strategies is that Eve learns everything for one Alice's setting and for the other an error between Alice and Bob occur in half of the measurements. Explicitly, for L_1^r and L_2^r, after the pseudo-sifting phase Bob's symbols are:

- $b(y = 0) = b(y = 1) = a$ when $x = 0$: There is no error between Alice and Bob and Eve has full information.
- $b(y = 0) \neq b(y = 1)$ when $x = 1$: There is error in half of the measurements between Alice and Bob and Eve has no information about b.

For L_3^r and L_4^r, the situation is reversed, Eve has full information in $x = 1$ and no knowledge on Bob's symbol when $x = 0$, where there is error between Alice and Bob half of the time.

The interest of this relationship is purely theoretical: it is the first evidence of an analogue of quantum mechanical uncertainty relations in a generic non-local theory.

5.3.2 One-way classical post-processing

General case

As already discussed in chapter 3, the bound for the length of the achievable secret key rate under individual attacks is the Csiszár-Körner (CK) bound [32].

In the study of this protocol the possible use of additional data T described in (3.20) was not explored, the pre-processing consisting just in a bit flip with probability q (That is for Bob, $b \to b+1$ with probability q).

Consequently, we have an estimate $r_{CK} \leq R_{CK}$ for the achievable secret key rate. Recalling that mutual information is $I(X:Y) = H(X) - H(X|Y)$, the estimate for the CK bound can be written as

$$\begin{aligned} r_{CK} &= \max_{B' \leftarrow B} \left[I(A:B') - I(B':E) \right] \\ &= \frac{1}{2} \sum_{x=0,1} \max_{B' \leftarrow B} \left[I(A:B'|x) - I(B':E|x) \right]. \end{aligned} \tag{5.17}$$

Let's sketch the computation explicitly. In Table 5.2, one reads for $p(a,b) = \frac{1}{2}P(a,b|x=0) + \frac{1}{2}P(a,b|x=1)$:

$$\begin{aligned} P(0,0) &= \tfrac{1}{2}\left(\tfrac{p_{NL}}{2} + p_1^0 + p_2^0 + \tfrac{p_3^0+p_4^0}{2}\right) + \tfrac{1}{2}\left(\tfrac{p_{NL}}{2} + p_4^1 + p_3^0 + \tfrac{p_1^0+p_2^1}{2}\right) = \\ &= \tfrac{1}{2} + \tfrac{p_1^0+p_3^0-p_2^1-p_4^0-2p_3^1-2p_1^1}{4} \\ P(0,1) &= \tfrac{p_1^0+p_2^1+p_3^0+p_4^0}{4} \\ P(1,0) &= \tfrac{p_1^1+p_2^0+p_3^1+p_4^1}{4} \\ P(1,1) &= \tfrac{1}{2}\left(\tfrac{p_{NL}}{2} + p_1^1 + p_2^1 + \tfrac{p_3^1+p_4^1}{2}\right) + \tfrac{1}{2}\left(\tfrac{p_{NL}}{2} + p_3^1 + p_4^0 + \tfrac{p_1^1+p_2^0}{2}\right) = \\ &= \tfrac{1}{2} + \tfrac{p_1^1+p_3^1-p_2^0-p_4^1-2p_3^0-2p_1^0}{4} \end{aligned} \tag{5.18}$$

We can also calculate explicitly the local probabilities for Alice and Bob.

$$\begin{aligned} P_A(0) &= \frac{1}{2} + \frac{p_1^0 + p_3^0 - p_1^1 - p_3^1}{2} \\ P_A(1) &= \frac{1}{2} + \frac{p_1^1 + p_3^1 - p_1^0 - p_3^0}{2} \end{aligned} \tag{5.19}$$

$$\begin{aligned} P_B(0) &= \frac{1}{2} + \frac{p_1^0 + p_2^0 + p_3^0 + p_4^1 - p_1^1 - p_2^1 - p_3^1 - p_4^0}{4} \\ P_B(1) &= \frac{1}{2} + \frac{p_1^1 + p_2^1 + p_3^1 + p_4^0 - p_1^0 - p_2^0 - p_3^0 - p_4^1}{4} \end{aligned} \tag{5.20}$$

We observe that the case of marginal where local probabilities are uniform is produced by the same probability of local strategies with $r = 0, 1$, i.e. $p_j^0 = p_j^1 = \frac{1}{2}$ $j = 1, 2, 3, 4$
These probabilities allow us to calculate the mutual information

$$\mathrm{I}(A;B) = \sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} P(a,b) \log \frac{P(a,b)}{P_A(a) P_B(b)} \tag{5.21}$$

Being q the probability that Bob flips his bit in the pre-processing, then the probability of each pair of symbols is:

$$P(a,b') = (1-q)P(a,b=b') + qP(a,b=b'+1). \tag{5.22}$$

These four probabilities allow to compute the mutual information in case of pre-processing $I(A;B') = H(A) - H(A|B')$.

Turning to Eve: before pre-processing, she has full knowledge on Bob's symbol for L_1^r and L_2^r, and no knowledge for L_3^r and L_4^r. Consequently

$$I(B' : E|x = 0) \quad = \quad I(B : E|x = 0) \left[1 - h(q)\right] \tag{5.23}$$

where h is binary entropy.

The calculation is of course identical for $x = 1$ and this allows to compute r_{CK} for any probability distribution. Equivalently one can express this as:

$$H(B'|E) \quad = \quad H(B|E) + \left[1 - H(B|E)\right] h(q) \tag{5.24}$$

In the following paragraphs, we will focus on two cases and will explicitly derive this calculations.

Isotropic distribution

Let's consider the probability distribution of the form

$$P(a, b|x, y) \quad = \quad \frac{1 + p_{NL}}{4} \, \delta(a + b = xy) + \frac{p_L}{8} \, . \tag{5.25}$$

This distribution is called an isotropic probability distribution. The forced symmetry in the distribution necessarily implies $p_j^r = p_L/8$ for all j, r, since the L_j^r are linearly independent. Note that the point of highest violation in the quantum region is of this form, with $p_{NL} = \sqrt{2} - 1 \simeq 0.414$ and probability distribution $P(a + b = xy|x, y) = \frac{1 + \frac{1}{\sqrt{2}}}{2}$.

Remarkably, Alice and Bob can transform any distribution with a given p_{NL} to the isotropic distribution defined by the same p_{NL} (or the same degree of violation of the CHSH inequality) with local operations and public communication. This procedure is called "depolarization" (which was introduced in [15]) and is implemented employing three bits of shared randomness and local operation with the following steps:

1. Alice and Bob perform one of the following operations with probability $\frac{1}{2}$

   - Do nothing
   - Flip a and b, that is $a = a + 1 \mod 2$ and $b = b + 1 \mod 2$

   This procedure unbias the correlations.

2. With probability $\frac{1}{4}$ they perform one of the following operations

   - Do nothing
   - $a \to a + x \mod 2$ and $y \to y + 1$
   - $b \to b + y \mod 2$ and $x \to x + 1$
   - $a \to a + x + 1 \mod 2$, $b \to b + y \mod 2$ and $x \to x + 1$

This implies that the results of this subsection are in some sense generic. In fact, by Theorem 2 of section 5.2, Eve's best individual eavesdropping strategy for a fixed value of p_{NL} consists in preparing an isotropic distribution. Alternatively, we can modify the protocol to add the fact that Alice and Bob apply systematically the depolarization procedure.

For isotropic distributions, the two tables for $x = 0$ and $x = 1$ become identical, and we can rewrite them as Table 5.3. In this Table, we have changed the notation for Eve's knowledge, and have written (a, b) when Eve knows both outcomes, $(a, ?)$ when she knows only Alice's, and $(?, ?)$ when she knows none.

This distribution has $p(a = 0) = p(a = 1) = \frac{1}{2}$. Before pre-processing, the error between Alice and Bob is $e_{AB} = p_L/4$; after pre-processing, the quantity to be corrected in error correction is $e'_{AB} = (1 - q)e_{AB} + q(1 - e_{AB})$. As stated in the previous section, we can use to obtain the

[isotropic]	$b=0$	$b=1$
$a=0$	$p_{NL}/2$ $(?,?)$ $p_L/4$ $(0,0)$ $p_L/8$ $(0,?)$	$p_L/8$ $(0,?)$
$a=1$	$p_L/8$ $(1,?)$	$p_{NL}/2$ $(?,?)$ $p_L/4$ $(1,1)$ $p_L/8$ $(1,?)$

Table 5.3: Probability distribution Alice-Bob-Eve for the CHSH protocol, in the case of isotropic distribution.

mutual information $\mathrm{I}(A;B)=\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}}P(a,b)\log\dfrac{P(a,b)}{P_A(a)P_B(b)}$. In the case of $\mathrm{I}(A;B')$ we have inspecting table 5.3

$$\begin{aligned}P(0,0) &= \frac{1}{2}-\frac{p_L}{8}\\ P(0,1) &= \frac{p_L}{8}\\ P(1,0) &= \frac{p_L}{8}\\ P(1,1) &= \frac{1}{2}-\frac{p_L}{8}\end{aligned}$$

And considering pre-processing:

$$\begin{aligned}P(0,0') &= (1-q)(\frac{1}{2}-\frac{p_L}{8})+q\frac{p_L}{8}\\ P(0,1') &= (1-q)\frac{p_L}{8}+q(\frac{1}{2}-\frac{p_L}{8})\\ P(1,0') &= (1-q)\frac{p_L}{8}+q(\frac{1}{2}-\frac{p_L}{8})\\ P(1,1') &= (1-q)(\frac{1}{2}-\frac{p_L}{8})+q\frac{p_L}{8}\end{aligned}$$

Thus, we have

$$\begin{aligned}\mathrm{I}(A;B') &= 2\left[(1-q)(\frac{1}{2}-\frac{p_L}{8})+q\frac{p_L}{8}\right]\log\left[\frac{(1-q)(\frac{1}{2}-\frac{p_L}{8})+q\frac{p_L}{8}}{\frac{1}{2}\cdot\frac{1}{2}}\right]+\\ &\quad +2\left[(1-q)\frac{p_L}{8}+q(\frac{1}{2}-\frac{p_L}{8})\right]\log\left[\frac{(1-q)\frac{p_L}{8}+q(\frac{1}{2}-\frac{p_L}{8})}{\frac{1}{2}\cdot\frac{1}{2}}\right]=\\ &= \left[(1-q)(1-\frac{p_L}{4})+q\frac{p_L}{4}\right]\log\left(2(1-q)(1-\frac{p_L}{4})+q\frac{p_L}{4}\right)+\\ &\quad +\left[(1-q)\frac{p_L}{4}+q(1-\frac{p_L}{4})\right]\log\left(2(1-q)\frac{p_L}{4}+q(1-\frac{p_L}{4})\right)=\\ &= 1-h(e'_{AB})\end{aligned}\tag{5.26}$$

Eve's information can be obtained as $I(B':E|x=0)=I(B:E|x=0)\left[1-h(q)\right]$. We compute explicitly the mutual information between Eve and Bob with the information in Table 5.3, for each symbol (b,e), we have the following terms in the mutual information formula (each term is $p(b,e)\log\frac{p(b,e)}{p_B(b)p_E(e)}$) (we explicitly show case for $b=0$)):

$$\begin{aligned}(0,(?,?)): &\quad \frac{p_{NL}}{2}\log\frac{\frac{p_{NL}}{2}}{\frac{1}{2}p_{NL}}=0\\ (0,(0,0)): &\quad \frac{p_L}{4}\log\frac{\frac{p_L}{4}}{\frac{1}{2}\frac{p_L}{4}}=\frac{p_L}{4}\\ (0,(0,?)): &\quad \frac{p_L}{8}\log\frac{\frac{p_L}{8}}{\frac{1}{2}\frac{p_L}{4}}=0\\ (0,(1,?)): &\quad \frac{p_L}{8}\log\frac{\frac{p_L}{8}}{\frac{1}{2}\frac{p_L}{4}}=0\end{aligned}$$

Therefore, Eve's mutual information in the isotropic case is $\frac{p_L}{2}[1-h(q)]$. We can now explicitly calculate the CK bound defined previously

$$r_{CK} = \max_{q\in[0,\frac{1}{2}]} \left[1-h(e'_{AB})-\frac{p_L}{2}[1-h(q)]\right]. \quad (5.27)$$

This quantity is plotted in Fig. 5.1 as a function of the local probability $p_L = 1-p_{NL}$. We see that $r_{CK} > 0$ for $p_L \leq 0.764$, i.e. $p_{NL} \gtrsim 0.236$ for the optimal pre-processing. Without pre-processing, the bound becomes $p_L \leq 0.682$, i.e. $p_{NL} \gtrsim 0.318$. The important remark is that both these values are inside the quantum region. This means that using quantum physics, one can distribute correlations which allow (at least against individual attacks) the extraction of a secret key without any further assumption about the details of the physical realization.

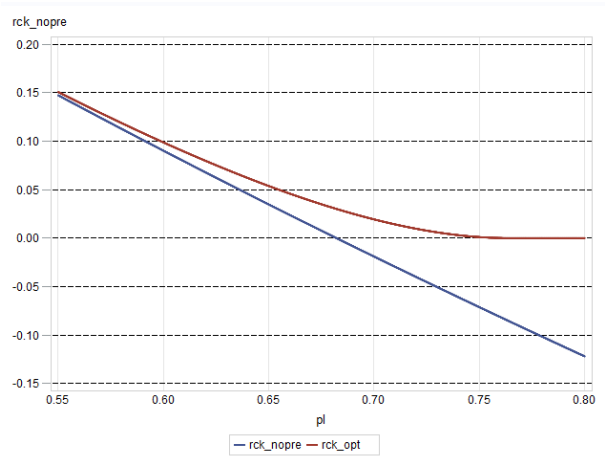


Figure 5.1: Achievable secret key rate for the CHSH protocol, after one-way post-processing against a no-signalling Eve for individual attacks, isotropic distribution

In the paper derived from this work [18], the CHSH protocol was analyzed with the standard approach of quantum cryptography. The details will not be exposed here (see Appendix A of [18]) but it is interesting to show the results. In this approach Alice and Bob share a quantum state of two qubits and have agreed on the physical measurements corresponding to each value of x and y; Eve is constrained to distribute quantum states, of which she keeps a purification.

The resulting bound on the achievable secret key rate is plotted in Fig. 5.2. This quantity is plotted as a function of the disturbance D defined by $p_{NL} = \sqrt{2}(1-2D)-1$. This parameter characterizes the properties of the channel linking Alice and Bob: it is therefore useful for comparison with a quantum realization of the CHSH protocol and with BB84. It turns out that the CHSH protocol is equivalent to the BB84 protocol plus some classical pre-processing. In particular, the robustness to noise is the same for both protocols. For low error rate, BB84 provides higher secret-key rate; however, BB84 cannot be used for a device-independent proof, since (as we noticed above) its correlations become intrinsically insecure if the dimensionality of the Hilbert space is not known.

Reaching the Bell limit

Another interesting case is to explore if one can find one-parameter families of probability distributions for which $R_{CK} > 0$ as soon as $p_{NL} > 0$. This means finding distributions for which one can extract a secret key out of one-way processing down to the limit of the local polytope. There are distributions of such type and even a secret key can be extracted even without pre-processing. Here is an example: set $p_1^0 = p_2^0$, $p_1^1 = p_2^1$, and $p_{3,4}^r = 0$. For both $x = 0$ and $x = 1$ we have $p(a = 0) = p(a = 1) = \frac{1}{2}$. For $x = 0$, Alice and Bob make no errors ($e_{AB|0} = 0$), and Eve's information is $I(B : E|x = 0) = p_L$; for $x = 1$, the errors of Alice and Bob are $e_{AB|1} = \frac{p_L}{2}$ and Eve has no information. In summary, even neglecting pre-processing,

$$r_{CK} = 1-\frac{1}{2}h(p_L/2)-\frac{p_L}{2} \quad (5.28)$$

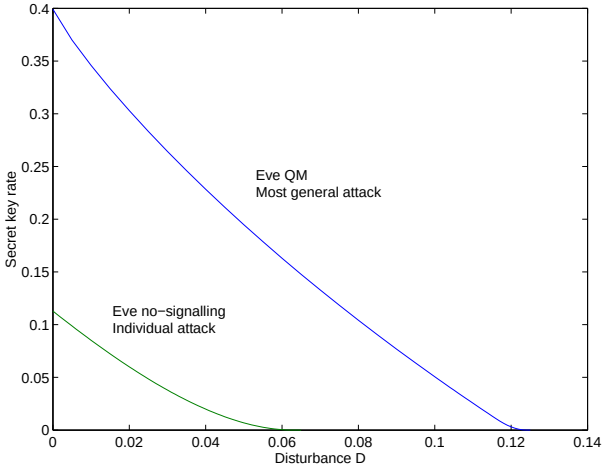


Figure 5.2: Achievable secret key rate for the CHSH protocol, after one-way post-processing: against a no-signalling Eve for individual attacks, isotropic distribution (5.3.2) and against a quantum Eve, in a two-qubit implementation

which is strictly positive in the whole region $p_L < 1$.

However, the distributions described here cannot be broadcasted using quantum states. The reason is that the quantum intersection with the non-local region is strictly inside this region: as soon as $p_{NL} > 0$, all the p_j^r must be non zero, because the L_j^r are linearly independent to consider a point in the quantum set. On the contrary, here we have set $p_{3,4}^r = 0$. Anyway, in spite of the fact that we are not able to broadcast this distribution with known physical means, it is interesting to notice that there exists a family of probability distributions that can lead to a secret key under one-way post-processing, for any amount of non-locality.

5.3.3 Two-way classical post-processing

Contrary to the one-way case, no tight bound like the Csiszár-Körner bound is known when two-way classical post-processing is allowed; nor is the optimal procedure known. The best-known two-way post-processing is advantage distillation, described in chapter 3. Forgetting about pre-processing, one can see the effect as follows: starting from a situation where $I(A:B) < I(B:E)$, one makes a processing at the end of which the new variables satisfy $I(\tilde{A}:\tilde{B}) > I(\tilde{B}:\tilde{E})$; at this point, one applies the one-way post-processing.

We apply the procedures described in 3.2.1 to the isotropic correlations described above (5.3.2), first without pre-processing, then by allowing Alice and Bob to perform some bit flip before starting the advantage distillation. We anticipate the result: we find that a key can be extracted for $p_{NL} \gtrsim 0.09$; that is, even with two way post-processing we are not able to reach the Bell limit for isotropic correlations.

Advantage distillation without pre-processing

We refer to Table 5.3. We have, as above, $e_{AB} = \frac{p_L}{4}$. We must now estimate Eve's error on Bob's symbol after AD. Eve knows α as soon as she knows one of Alice's symbols a_{i_k}, and recall that asymptotically the guess $\beta = \alpha$ is correct. The only situation in which Eve is obliged to make a random guess is therefore the case in which all the N instances correspond to Eve's symbol $(?,?)$. The probability that Eve's guess of Bob's symbol is wrong is therefore

$$\tilde{e}_E \gtrsim \frac{1}{2}\left(\frac{p_{NL}}{1-e_{AB}}\right)^N \tag{5.29}$$

where the denominator comes from the fact that we must condition on the bit's acceptance. Using (3.17), we obtain that secrecy can be extracted as long as $p_{NL} > p_L/4$ that $p_{NL} > 1/5$. This is lower than the bound obtained for one-way post-processing, as expected.

Advantage distillation with pre-processing

The previous bound can be further improved by allowing Alice and Bob to pre-process their lists of symbols before starting the advantage distillation procedure.

For two-way post-processing, it is not known whether bitwise pre-processing is already optimal; but we restricted to it in this work.

Let's analyze in first instance, one way pre-processing, an then generalize the results to two-way pre processing. This result is based on the paper [3] where a bound for advantage distillation is derived for the two-qubit entanglement case.
If Alice performs some pre-processing, where she flips her measurement with probability q_A, otherwise she keeps it (with probability $\bar{q}_A = 1 - q_A$), the resulting probability distribution becomes

	$b=0$	$b=1$
$a=0$	$\bar{q}_A \frac{p_{NL}}{2}$ $(?,?)$ $\bar{q}_A \frac{p_L}{4}$ $(0,0)$ $\bar{q}_A \frac{p_L}{8}$ $(0,?)$ $q_A \frac{p_L}{8}$ $(1,?)$	$q_A \frac{p_{NL}}{2}$ $(?,?)$ $q_A \frac{p_L}{4}$ $(1,1)$ $\frac{p_L}{8}\bar{q}_A$ $(0,?)$ $q_A \frac{p_L}{8}$ $(1,?)$
$a=1$	$q_A \frac{p_{NL}}{2}$ $(?,?)$ $q_A \frac{p_L}{4}$ $(0,0)$ $q_A \frac{p_L}{8}$ $(0,?)$ $\bar{q}_A \frac{p_L}{8}$ $(1,?)$	$\bar{q}_A \frac{p_{NL}}{2}$ $(?,?)$ $q_A \frac{p_L}{4}$ $(1,1)$ $q_A \frac{p_L}{8}$ $(0,?)$ $\bar{q}_A \frac{p_L}{8}$ $(1,?)$

Table 5.4: Probability distribution Alice-Bob-Eve for the sifted data, in the case of isotropic distribution, after Alice's pre-processing.

Since the initial error probability was $e_{AB} = \frac{p_L}{4}$, with preprocessing, the error probability becomes $e_{A'B} = \bar{q}_A p_L/4 + q_A(1 - p_L/4)$. After advantage distillation with a block of N bits, the resulting error probability satisfies, as already stated

$$\tilde{e}_{A'B} \gtrsim \left(\frac{e_{A'B}}{1 - e_{A'B}}\right)^N \tag{5.30}$$

Concerning Eve's error, we can restrict the considerations to the cases where there is no error between Alice and Bob. As already assumed, as soon Eve guesses correctly Alice's symbol (α), she guesses Bob's (β). The question here is to derive a condition stronger than (5.29), having into consideration the uncertainty introduced by Alice's pre-processing in Eve's knowledge. Eve's situation is such that, even if she has a symbol (a, b) or $(a, ?)$ she cannot be completely sure of Alice's symbol. In order to bound her error probability, we can restrict our considerations to the cases where $a = b$, since $e_{A'B} \to 0$.

The relevant terms to focus on in order to bound Eve's error are for instance the probability that Eve has the symbol $(0, 0)$ and the bits are accepted

$$P_{eq}(0,0|0) = P(e = (0,0)|a = 0, b = 0) = \frac{\bar{q}_A p_L}{4\nu_q} \tag{5.31}$$

$$\tag{5.32}$$

where ν_q is the probability of acceptance of the bit, that is $\nu_q = \frac{1-e_{A'B}}{2}$.

$$\nu_q = P(a = 0, b = 0) = P(a = 1, b = 1) = \frac{p_L}{8} + \bar{q}_A(\frac{p_L}{4} + \frac{1 - p_L}{2}) \tag{5.33}$$

In order to bound Eve's error probability after advantage distillation, one concentrates in instances where

1. The number of zeros and ones in the vector $\vec{x}$ used for advantage distillation is the same
2. No $(0, 0)$ or $(1, 1)$ appears and therefore Eve has no deterministic information.
3. In all the positions where Eve expects to have the same symbol, the number of $(0, ?)$ and $(1, ?)$ is the same.

One can derive the following expression

$$\tilde{e}_E \gtrsim \frac{1}{2}\frac{1}{2^N} \sum_{n_{NL0},n_{NL1},n_{L0},n_{L1}} \frac{N!}{n_{NL0}!n_{NL1}!(n_{L0}!)^2(n_{L1}!)^2} P_{eq}(?,?|0)^{n_{NL0}} P_{eq}(?,?|1)^{n_{NL1}} \\ P_{eq}(0,?|0)^{n_{L0}} P_{eq}(1,?|0)^{n_{L0}} P_{eq}(0,?|1)^{n_{L1}} P_{eq}(1,?|1)^{n_{L1}} \tag{5.34}$$

Where n_{NL0} and n_{NL1} denote the number of terms where the term $(?,?)$ appears in the zeros and ones of the vector $\vec{a}$. In the zeros (ones) of $\vec{a}$, Eve has n_{L0} (n_{L1}) terms of (0,?) and (1,?). So, we have $n_{NL0}+n_{NL1}+2(n_{L0}+n_{L1}) = N$. In these cases, Eve has no information about the symbol used for advantage distillation and she has to guess.
We explicitly express the previous expression, considering that:

$$P_{eq}((?,?)|0) = P_{eq}((?,?)|1) = \frac{\bar{q}_A(1-p_L)}{2\nu_q}$$
$$P_{eq}((0,?)|0) = P_{eq}((1,?)|1) = \frac{\bar{q}_A p_L}{8\nu_q}$$
$$P_{eq}((0,?)|1) = P_{eq}((1,?)|0) = \frac{q_A p_L}{8\nu_q}$$

Therefore,

$$\tilde{e}_E \gtrsim \frac{1}{2}\frac{1}{2^N} \sum_{n_{NL0},n_{NL1},n_{L0},n_{L1}} \frac{N!}{n_{NL0}!n_{NL1}!(2n_{L0})!(2n_{L1})!} \left(\frac{\bar{q}_A(1-p_L)}{2\nu_q}\right)^{n_{NL0}} \left(\frac{\bar{q}_A(1-p_L)}{2\nu_q}\right)^{n_{NL1}} \\ \binom{2n_{L0}}{n_{L0}} \left(\frac{\bar{q}_A p_L}{8\nu_q}\right)^{n_{L0}} \left(\frac{q_A p_L}{8\nu_q}\right)^{n_{L0}} \binom{2n_{L1}}{n_{L1}} \left(\frac{\bar{q}_A p_L}{8\nu_q}\right)^{n_{L1}} \left(\frac{q_A p_L}{8\nu_q}\right)^{n_{L1}} \tag{5.35}$$

The term $\frac{1}{2^N}$ appears because we are considering all the possible vectors of size N. Employing that for large n, $\frac{(2n)!}{(n!)} \simeq 2^{2n}$, we have:

$$\tilde{e}_E \gtrsim \frac{1}{2} \sum_{n_{NL0},n_{NL1},n_{L0},n_{L1}} \frac{N!}{n_{NL0}!n_{NL1}!(2n_{L0})!(2n_{L1})!} \left(\frac{\bar{q}_A(1-p_L)}{4\nu_q}\right)^{n_{NL0}} \left(\frac{\bar{q}_A(1-p_L)}{4\nu_q}\right)^{n_{NL1}} \\ \left(\frac{\sqrt{\bar{q}_A q_A}p_L}{8\nu_q}\right)^{2n_{L0}} \left(\frac{\sqrt{\bar{q}_A q_A}p_L}{8\nu_q}\right)^{2n_{L1}} \tag{5.36}$$

In the limit, as N increases, employing a multinomial expansion in even values, this expression can be approximated by:

$$\tilde{e}_E \gtrsim \frac{1}{2}\frac{1}{4}\left(\frac{\bar{q}_A(1-p_L)}{2\nu_q} + \frac{\sqrt{\bar{q}_A q_A}p_L}{4\nu_q}\right)^N \tag{5.37}$$

Combining this expression with equation (5.30), we have that the key distillation is possible whenever:

$$\frac{e_{A'B}}{1-e_{A'B}} < \frac{\bar{q}_A(1-p_L)}{2\nu_q} + \frac{\sqrt{\bar{q}_A q_A}p_L}{4\nu_q} \tag{5.38}$$

If $q_A = 0$, one obtains $p_L < 0.8$ as was already stated in the previous section. If we optimize $\bar{q}_A$ as a function of p_L, we obtain the critical value $p_L \simeq 0.845294$ to obtain a key[1].

[1]In order to optimize this quantity, one finds that the optimal q_A for these expressions is of the form:

$$\hat{q}_A = \frac{10p_L^2 \pm \sqrt{2}\sqrt{45p_L^4 - 228p_L^3 + 440p_L^2 - 384p_L + 128} - 24p_L + 16}{4(5p_L^2 - 12p_L + 8)} \tag{5.39}$$

Let's develop the two-way pre processing, changing the notation employed but with the same technique in order to obtain a bound for Eve's error and thus the condition to distill a secret key. We suppose that Alice flips her bit with probability q_A, Bob with probability q_B. By inspection, one finds that the probability distribution obtained from Table 5.3 after this two-way pre-processing is the one of Table 5.5.

	$b=0$	$b=1$
$a=0$	$\frac{p_{NL}}{2}(\bar{q}_A\bar{q}_B + q_Aq_B)$ $(?,?)$ $\frac{p_L}{4}\bar{q}_A\bar{q}_B$ $(0,0)$ $\frac{p_L}{4}q_Aq_B$ $(1,1)$ $\frac{p_L}{8}\bar{q}_A$ $(0,?)$ $\frac{p_L}{8}q_A$ $(1,?)$	$\frac{p_{NL}}{2}(q_A\bar{q}_B + \bar{q}_Aq_B)$ $(?,?)$ $\frac{p_L}{4}\bar{q}_Aq_B$ $(0,0)$ $\frac{p_L}{4}q_A\bar{q}_B$ $(1,1)$ $\frac{p_L}{8}\bar{q}_A$ $(0,?)$ $\frac{p_L}{8}q_A$ $(1,?)$
$a=1$	$\frac{p_{NL}}{2}(q_A\bar{q}_B + \bar{q}_Aq_B)$ $(?,?)$ $\frac{p_L}{4}q_A\bar{q}_B$ $(0,0)$ $\frac{p_L}{4}\bar{q}_Aq_B$ $(1,1)$ $\frac{p_L}{8}q_A$ $(0,?)$ $\frac{p_L}{8}\bar{q}_A$ $(1,?)$	$\frac{p_{NL}}{2}(\bar{q}_A\bar{q}_B + q_Aq_B)$ $(?,?)$ $\frac{p_L}{4}q_Aq_B$ $(0,0)$ $\frac{p_L}{4}\bar{q}_A\bar{q}_B$ $(1,1)$ $\frac{p_L}{8}q_A$ $(0,?)$ $\frac{p_L}{8}\bar{q}_A$ $(1,?)$

Table 5.5: Probability distribution Alice-Bob-Eve for the sifted data, in the case of isotropic distribution, after Alice's and Bob's pre-processing.

As in the case of pre-processing of Alice, just by looking at the Table, one can guess the interest of pre-processing: the five possible symbols for Eve are now spread in all the four cells of the table. For instance, Eve's symbol is $(0,0)$ was present only in the case $a=b=0$ in Table 5.3, that is, whenever she had this symbol Eve had full information; this is no longer the case in Table 5.5. Note also that the roles of q_A and q_B are not symmetric, because only q_A mixes the strategies for which Eve does not know Bob's symbol.

The distribution of Table 5.5 is such that

$$e'_{AB} \;=\; \left(p_{NL} + \frac{p_L}{2}\right)(q_A\bar{q}_B + \bar{q}_Aq_B) + \frac{p_L}{4}\,. \tag{5.40}$$

The estimate of Eve's error is derived in a similar way as in the case where only Alice flipped her symbol. Eve's situation now is such that, even if she has a symbol (a,b) or $(a,?)$, she cannot be completely sure whether $\alpha = a$ or not.

Suppose that among her N symbols, modifying the previous notation, Eve has n_0 times the symbol $(?,?)$, n_1^0 times the symbol $(0,?)$, n_1^1 times the symbol $(1,?)$, n_2^0 times the symbol $(0,0)$, and n_2^1 times the symbol $(1,1)$. Eve cannot avoid errors when $p(a=0|e) = p(a=1|e)$, that is when $n_1^0 = n_1^1 \equiv n_1$ and $n_2^0 = n_2^1 \equiv n_2$. We have therefore the bound

$$\tilde{e}'_E \;\gtrsim\; \frac{1}{2}\sum_{n_0,n_1,n_2} \frac{N!}{n_0!(n_1!)^2(n_2!)^2}\,\gamma_{(?,?)}^{n_0}\,\gamma_1^{2n_1}\,\gamma_2^{2n_2} \tag{5.41}$$

where the sum is taken under the constraint $n_0 + 2n_1 + 2n_2 = N$, γ_e is the probability that Eve has symbol e conditioned on the bit's acceptance, that is $\gamma_e = P_{eq}(e|0) + P_{eq}(e|1)$ in the previous notation and $\gamma_1 \equiv \sqrt{\gamma_{(0,?)}\gamma_{(1,?)}}$, $\gamma_2 \equiv \sqrt{\gamma_{(0,0)}\gamma_{(1,1)}}$. By using as before $(n!)^2 \sim (2n)!/2^{2n}$ and summing the multinomial expansion, we obtain

$$\tilde{e}'_E \;\gtrsim\; \frac{1}{8}\left(\gamma_{(?,?)} + 2\gamma_1 + 2\gamma_2\right)^N\,. \tag{5.42}$$

The expressions for the γ_e are derived from Table 5.5. Suppose for definiteness that Alice and Bob have accepted the bit $\alpha = \beta = 0$: this happens with probability $\frac{1-e'_{AB}}{2}$. The probability that this happens and that Eve has got the symbol $(?,?)$ is $\frac{p_{NL}}{2}(\bar{q}_A\bar{q}_B + q_Aq_B)$; whence $\gamma_{(?,?)} = \frac{p_{NL}(\bar{q}_A\bar{q}_B + q_Aq_B)}{1-e'_{AB}}$. Similarly, the probability that Alice and Bob accept the bit 0 and that Eve has got $(0,?)$, respectively $(1,?)$, is $\frac{p_L}{8}\bar{q}_A$, respectively $\frac{p_L}{8}q_A$; whence by symmetry respect bit 1, $\gamma_1 = \frac{p_L\sqrt{\bar{q}_Aq_A}}{4\,(1-e'_{AB})}$. In a similar way, γ_2 is computed having in consideration that probability of acceptance of bit 0 with Eve having $(0,0)$ is $\frac{p_L}{4}\bar{q}_A\bar{q}_B$ and Eve having $(1,1)$ is $\frac{p_L}{4}q_Aq_B$.

To simplify calculations we write $\delta_e \equiv (1 - e'_{AB})\gamma_e$, we have then

$$\begin{array}{lcl} \delta_{(?,?)} & = & p_{NL}\left(\bar{q}_A\bar{q}_B + q_A q_B\right), \\ \delta_1 & = & \frac{p_L}{4}\sqrt{\bar{q}_A q_A}\,, \\ \delta_2 & = & \frac{p_L}{2}\sqrt{\bar{q}_A q_A}\sqrt{\bar{q}_B q_B} \end{array} \tag{5.43}$$

and the condition for extraction of a secret key becomes

$$\delta_{(?,?)} + 2\delta_1 + 2\delta_2 \quad > \quad e'_{AB}\,. \tag{5.44}$$

The optimization over q_A and q_B can be done numerically. The result is that a secret key can be extracted at least down to $p_{NL} \approx 0.0939$, that is $p_L \leq 0.9061$.

Intrinsic information

We have already seen in chapter 2 the properties of intrinsic information and that for finite alphabets, the optimization can be performed explicitly (as stated in 3.4). We are now able to consider its calculation for the case of the isotropic probability distribution of the protocol. If we calculate mutual conditioned information in the case of the original probability distribution (table 5.3) we obtain that $\mathrm{I}(A;B \mid E) = p_{NL}$. The idea is to obtain a channel $E \to \overline{E}$ which minimizes this quantity, and prove that this quantity is always positive, given that this is a necessary condition for a key be stablished. Let's define the stochastic matrix of the channel:

$$P_{E|\overline{E}}(i,j) = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} \\ a_{10} & a_{11} & a_{12} & a_{13} & a_{14} \\ a_{20} & a_{21} & a_{22} & a_{23} & a_{24} \\ a_{30} & a_{31} & a_{32} & a_{33} & a_{34} \\ a_{40} & a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \tag{5.45}$$

Where entries are considered such that $0 = (0,0),\ 1 = (0,?),\ 2 = (?,?),\ 3 = (1,?),\ 4 = (1,1)$. The intrinsic information is obtained minimizing mutual conditioned information under this map:

$$\mathrm{I}(A;B \downarrow E) = \min_{a_{ij}} \mathrm{I}(A;B \mid \overline{E}) \tag{5.46}$$

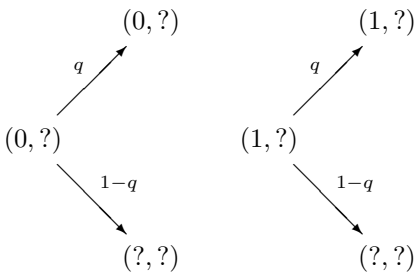
One can see that mutual conditioned information can be expressed in terms of entropies:

$$\begin{aligned} \mathrm{I}(A;B \mid E) \quad &= H(A \mid E) - H(A,B \mid E) = (H(A,E) - H(E)) - (H(A,B,E) - H(B,E)) = \\ &= H(A,E) + H(B,E) - H(A,B,E) - H(E) \end{aligned}$$

Thus, we have to minimize the function:

$$\begin{aligned} \mathrm{I}(A;B \downarrow E) \quad &= \min_{a_{ij}}(-\textstyle\sum_{a,\bar{e}} p(a,\bar{e})\log(p(a,\bar{e})) - \sum_{b,\bar{e}} p(b,\bar{e})\log(p(b,\bar{e})) + \\ &\quad + \textstyle\sum_{a,b,\bar{e}} p(a,b,\bar{e})\log(p(a,b,\bar{e})) + \sum_{\bar{e}} p(\bar{e})\log(p(\bar{e}))) \qquad (5.47) \\ &\quad \text{s.t.} \textstyle\sum_i a_{ij} = 1\ \forall\ j \qquad (5.48) \end{aligned}$$

Where $p(z,\bar{e}) = \sum_e p(z,e)p_{E|\overline{E}}(\bar{e},e)$ being $z = \emptyset,\ a,\ b,\ (a,b)$. We have done this optimization numerically, and found out that the channel which gives us the intrinsic information is of the form:



with $q = 1$. The choice of this map is justified intuitively in terms of *noising* nonlocal correlations. Considering the map, the stochastic matrix is of the form

$$P_{E|\overline{E}}(i,j) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & q & 0 & 0 & 0 \\ 0 & 1-q & 1 & 1-q & 0 \\ 0 & 0 & 0 & q & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.49}$$

defined like this trying to minimize the effect of (?,?)(We have that $I(A:B \mid E) = p_{NL}$ with the original distribution). This map looks conter-intuitive from Eve's point of view. Why should she map $(0,?)$ and $(1,?)$, where she knows Alice's variable, into $(?,?)$? Recall however that this map is only a mathematical tool for the computation of an upper bound on the secret-key rate. The new probability distribution is

b a (e)	0	1
0	(0,0) $p_L/4$ (?,0) $q \cdot p_L/8$ (?,?) $p_{NL}/2 + (1-q) \cdot p_L/8$	 (?,0) $q \cdot p_L/8$ (?,?) $(1-q) \cdot p_L/8$
1	 (?,1) $q \cdot p_L/8$ (?,?) $(1-q) \cdot p_L/8$	(1,1) $p_L/4$ (?,1) $q \cdot p_L/8$ (?,?) $p_{NL}/2 + (1-q) \cdot p_L/8$

The mutual conditioned information is then:

$$\begin{aligned} I(A:B \mid \overline{E})_q = \; & (1-(1+q)\tfrac{p_L}{2})(1+\tfrac{1/2-(3+q)p_L/8}{1/2-(1+q)p_L/4}\log\left(\tfrac{1/2-(3+q)p_L/8}{1/2-(1+q)p_L/4}\right) + \\ & \tfrac{(1-q)p_L/8}{1/2-(1+q)p_L/4}\log\left(\tfrac{(1-q)p_L/8}{1/2-(1+q)p_L/4}\right)) \\ = \; & (1-(1+q)\tfrac{p_L}{2})(1-h(\tfrac{1/2-(3+q)p_L/8}{1/2-(1+q)p_L/4})) \end{aligned} \tag{5.50}$$

This function of q is increasing -it can be seen numerically, or from the fact that $\forall p_L \in (0,1)$ $\frac{\partial}{\partial q}(1-(1+q)\frac{p_L}{2})(1-h(\frac{1/2-(3+q)p_L/8}{1/2-(1+q)p_L/4}))$ is positive. So the minimum of $I(A:B \mid \overline{E})_q$ is obtained for $q = 1$. This is a good candidate for the intrinsic information $\mathrm{I}(A;B \downarrow E)$.
If we consider the cases $(0,0) \to (0,0)$ with probability q and $(0,0) \to (?,?)$ with probability $1-q$, (and the symmetric with (1,1) simultaneously) we've seen numerically that it doesn't optimize the present result. We can combine this results and find that the present mutual conditioned information is the minimum in the combination of channels of the form $(x,x) \to^{q} (x,x)$ and $(x,x) \to^{1-q} (?,?)$ $E \to \overline{E} \to \overline{\overline{E}}$.
This candidate for intrinsic information is always positive. The mutual conditioned information under this map is given by

$$\mathrm{I}(A;B \mid \overline{E}) = (1-\tfrac{p_L}{2})(1+\tfrac{1/2-3p_L/8}{1/2-p_L/4}\log\left(\tfrac{1/2-3p_L/8}{1/2-p_L/4}\right) + \tfrac{p_L/8}{1/2-p_L/4}\log\left(\tfrac{p_L/8}{1/2-p_L/4}\right)) \tag{5.51}$$

$$= (1-\tfrac{p_L}{2})(1-h\left(\tfrac{p_L}{4-2p_L}\right)) \tag{5.52}$$

Interestingly, the quantity $\mathrm{I}(A;B \mid E)$ is positive whenever $p_{NL} > 0$. If the conjecture is true, it implies that either (i) it is possible to have a positive secret-key rate for the whole region of Bell violation, using a new key-distillation protocol, or (ii) the probability distribution of Table 5.2 represents an example of bipartite bound information for sufficiently small values of p_{NL}.

5.4 Protocol for Qudits

In this section we will continue the description from the previous chapter of the non-local polytope for $d > 2$, describe a generalization of the CHSH protocol and study the best strategies from Eve's

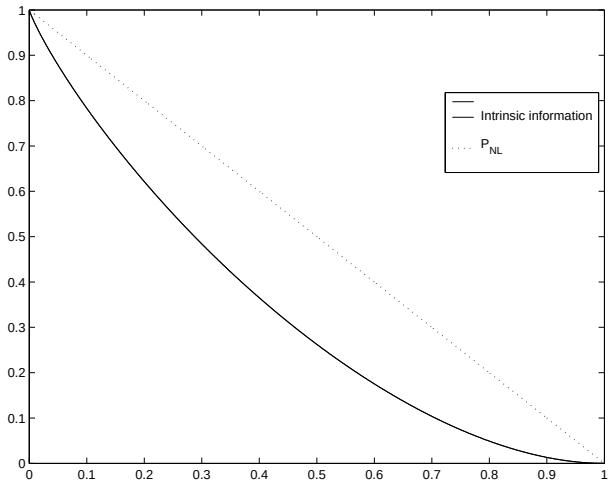


Figure 5.3: Representation of Intrinsic information for $d = 2$ as a function of p_L

viewpoint. Then, we will analyze in detail secrecy extraction in the case $d = 3$ following the same steps as above, that is one-way post-processing, two-ways post-processing and study of the intrinsic information measure.

We recall the slice described in 4.2.4, where the probability distributions associated to the optimal settings belong to a very symmetric family. Specifically, these distributions are such that

(i) for fixed inputs x and y, $P(a, b|x, y)$ depends only on $\Delta = a - b$; and

(ii) the probabilities for the different inputs are related as $P(\Delta|0, 0) = P(-\Delta|0, 1) = P(-\Delta|1, 0) = P(\Delta - 1|1, 1)$.

Compactly:

$$P(a, b = a - \Delta|x, y) \quad = \quad \frac{1}{d} p_f \tag{5.53}$$

with $f = (-1)^{x+y}\Delta + xy$ and $p_f = \sum_a P(a, b = a - f|0, 0)$.

As it happened for the isotropic distributions for $d = 2$, there exists a depolarization procedure that maps any probability distribution onto this slice by local operations and public communication, while keeping the violation $\langle I_d, P\rangle$ constant. As a consequence, Eve's optimal individual eavesdropping, for a fixed value of the violation of the inequality, consists in distributing a point in the described slice. This protocol can be realized with $\lceil \log(d) \rceil + 2$ bits of communication.

The procedure is:

- First of all we must uniformize locally Alice and Bob's inputs. This is made by:
  1. With probability $\frac{1}{d}$: $a \to a$ and $b \to b$ (That is do nothing!)
  2. With probability $\frac{1}{d}$: $a \to a + 1$ and $b \to b + 1$
  3. $\vdots$
  4. With probability $\frac{1}{d}$: $a \to a + (d - 1)$ and $b \to b + (d - 1)$

  This implements $P \to P_1$ which is such that $P_1(a, b|x, y) = \frac{1}{d} \sum_k P(a + k, b + k|x, y)$ and is consequently a function only of $\Delta = a - b$.

- Afterwards, we must obtain the probability distribution doing some operation on inputs and outputs (we thought in first term to use some operations which conservated the nonlocal machine $PR_{2,d}$, but finally the transformations were different...)
  1. With probability $\frac{1}{4}$: Do nothing

2. With probability $\frac{1}{4}$:

Flip x	y
$a \to -a$	$b \to -b+y$

3. With probability $\frac{1}{4}$:

x	Flip y
$a \to -a-x$	$b \to -b$

4. With probability $\frac{1}{4}$:

Flip x	Flip y
$a \to -a+x$	$b \to -b+\bar{y}$

where $\bar{0}=1$ and $\bar{1}=0$.
This implements $P_1 \longrightarrow P_2$ such that

$$\begin{aligned}
4P_2(a,b|0,0) &= P_1(a,b|0,0)+P_1(-a,-b|0,1)\\
&\quad +P_1(-a,-b|1,0)+P_1(a,b+1|1,1)\\
4P_2(a,b|0,1) &= P_1(a,b|0,1)+P_1(-a,-b+1|1,1)\\
&\quad +P_1(-a,-b|0,0)+P_1(a,b|1,0)\\
4P_2(a,b|1,0) &= P_1(a,b|1,0)+P_1(-a-1,-b|1,1)\\
&\quad +P_1(-a,-b|0,0)+P_1(a+1,b+1|0,1)\\
4P_2(a,b|1,1) &= P_1(a,b|1,1)+P_1(-a,-b+1|0,1)\\
&\quad +P_1(-a-1,-b|1,0)+P_1(a+1,b|0,0)\,.
\end{aligned}$$

Because of the symmetry of P_1, this implies $P_2(a,b|0,0)=P_2(-a,-b|0,1)=P_2(-a,-b|1,0)=P_2(a,b+1|1,1)$ which is nothing but the definition of the slice (5.53)

For example, none of the extremal points of the form $PR_{2,d'}$, with $d'<d$, is on the slice. Let's then consider a realization of $PR_{2,d'}$, the one whose array is

$$\hat{P}(d') = \frac{1}{d'}\begin{array}{|c|cc|cc|}\hline A\setminus B & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0}\\ \hline \mathbf{1} & \mathbb{1}_{d'} & \mathbf{0} & \mathbb{1}_{d'} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \hline \mathbf{1} & \mathbb{1}_{d'} & \mathbf{0} & U_{d'} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \hline\end{array} \tag{5.54}$$

where boldface numbers indicate arrays containing all ones or all zeros, $\mathbb{1}_{d'}$ is the identity matrix of dimension $d'\times d'$, and where $U_{d'}$ is the $d'\times d'$ matrix

$$U_{d'} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0\\ 0 & 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 0 & 1 & & 0\\ \vdots & \vdots & & & \ddots & \\ 0 & 0 & 0 & 0 & \cdots & 1\\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{5.55}$$

The arrays (4.29) and (5.54) allow to compute immediately the "scalar product"

$$\left\langle I_d, \hat{P}(d')\right\rangle = 1-\frac{1}{d'} \tag{5.56}$$

generalizing the results we gave in the main text for $d'=2$ and $d'=d$.

By following the steps of the depolarization protocol, one finds that $\hat{P}(d')$ goes to the distribution in the slice which is given by

$$\hat{P}(d') \longrightarrow \hat{P}_2(d') \equiv \left\{\begin{array}{lcl} p_0 &=& 1-\frac{1}{4d'}\\ p_{d'} &=& \frac{1}{4d'}\end{array}\right. \tag{5.57}$$

and obviously all the other p_f are zero. Using (5.64), one can verify that $\left\langle I_d, \hat{P}_2(d')\right\rangle = 1-\frac{1}{d'}$: the violation is preserved.
In general, the depolarization preserves the violation of the CGLMP inequality, that is

$$\langle I_d, P_2\rangle = \langle I_d, P\rangle\,. \tag{5.58}$$

The easiest way is to write down I_d as it appeared in the original paper [21], namely $\tilde{I}_d \leq 2$ with

$$\begin{aligned}\tilde{I}_d &= \sum_{k=0}^{[d/2]-1}\left(1-\frac{2k}{d-1}\right)\Big\{\big[P(-k|0,0)+P(k|0,1)\\ &\quad +P(k|1,0)+P(-k-1|1,1)\big]-\big[P(k+1|0,0)\\ &\quad +P(-k-1|0,1)+P(-k-1|1,0)+P(k|1,1)\big]\Big\}\end{aligned}$$

where $P(\Delta|x,y)\equiv P(a-b=\Delta|x,y)$. Using the expression of $\tilde{I}_d$, the statement is clear. In fact, Step 1 keeps by definition all the $P(\Delta|x,y)$ constant, while Step 2 keeps both sums in [...] constant.

Thus, the generalized depolarization procedure is an analog of the one applied to obtain the isotropic distribution. As in that case, we will be able to study a cryptographic protocol in a more simplified way.

5.4.1 Cryptography

The protocol

We suppose from the beginning $p(x=i)=p(y=j)=\frac{1}{2}$. The protocol is the analog of the CHSH protocol described above. When Alice announces $x=1$ and Bob has measured $y=1$, Bob corrects his dit according to $b\to b-1$. In other words, the pseudo-sifting implements $\Delta\to\Delta-xy$. The Alice-Bob distribution after pseudo-sifting, averaged on Bob's settings, becomes independent of x (as in the case of isotropic distribution for $d=2$):

$$P(a,a-\Delta|x) \;=\; \frac{1}{d}\sum_{y=0,1}p_{(-1)^{x+y}\Delta} \;=\; \frac{p_\Delta+p_{-\Delta}}{2d}\,. \tag{5.59}$$

In a protocol with d-dimensional outcomes, Alice and Bob can estimate not just one, but several error rates, one for each value of Δ. From the previous expression, we have that these error rates exhibit the symmetry

$$e_{AB}(\Delta) \;=\; e_{AB}(-\Delta) \;=\; \frac{p_\Delta+p_{-\Delta}}{2}\,. \tag{5.60}$$

As in the case of bits, we think of Eve as sending either a local or a non-local probability distribution. Let's discuss in some detail the points which lie on and above a CGLMP facet.

Eve's strategy: local points

In what follows, a characterization of the deterministic strategies that saturate the CGLMP inequality is detailed. It is not a full characterization as in the case of binary outcomes, but some facts are necessary for the full characterization of the case $d=3$.

Consider the array which represents the CGLMP inequality $I_d\leq 0$, eq. (4.29). It is more convenient to analyze the deterministic strategies to look at it as having $2d\times 2d$ entries. The array has the same form as the added terms cancel out. Let $I[i,j]$ denote an entry of this array. For the deterministic strategy $\big\{a(0),a(1);b(0),b(1)\big\}$, the value of CGLMP can be written as:

$$\begin{aligned}I_d &= -2+\sum_{x,y=0}^{1} I[a(x),b(y)] =\\ &= -2+\delta[b(0)\geq a(0)]+\delta[a(0)\geq b(1)]+\delta[a(1)\geq b(0)]-\delta[a(1)\geq b(1)]\end{aligned} \tag{5.61}$$

where the -2 comes from the marginals of $a(0)$ and $b(0)$, and where $\delta[C]$ is Kronecker's delta. The inequality is saturated by all the strategies such that $I_d=0$.

The first fact to be noted is that, for $d>2$, the number of deterministic points on the CGLMP facet is strictly larger than $D=4d(d-1)$, the dimension of the local and the no-signalling polytope.

This implies that, for some points on the facet, more than one decomposition as a convex combination of extremal points are possible. To prove this we consider the four "natural" relations associated to the CGLMP inequality, those that are simultaneously fulfilled by $PR_{2,d}$,:

$$\begin{array}{lcl} R_{00} & : & a(0) = b(0) \\ R_{01} & : & a(0) = b(1) \\ R_{10} & : & a(1) = b(0) \\ R_{11} & : & a(1) = b(1) - 1\,. \end{array} \tag{5.62}$$

It can indeed be verified using (5.61) that all the points that fulfill at least two among these relations saturate the inequality. There are $4d$ strategies that fulfill three relations, that are $\{R_{00}, R_{01}, R_{10}\}$, $\{R_{00}, R_{01}, R_{11}\}$, $\{R_{00}, R_{10}, R_{11}\}$, $\{R_{01}, R_{10}, R_{11}\}$ for each value of d.

The number of strategies that fulfill exactly two relations is obtained from the groupings

$$\begin{array}{lll} \{R_{00}, R_{01}\} & \{R_{00}, R_{10}\} & \{R_{00}, R_{11}\} \\ \{R_{01}, R_{10}\} & \{R_{01}, R_{11}\} & \{R_{10}, R_{11}\} \end{array} \tag{5.63}$$

in which the other two relationships are not fulfilled. That is, for example if $a(0) = b(0)$ and $a(0) = b(1)$ then $a(1) \neq b(0)$ and $a(1) \neq b(1) - 1$. This means that there are $6d(d-2)$ such strategies. Therefore, the number of strategies that fulfill at least two of the four relations in (5.62) is $6d(d-2) + 4d = 6d^2 - 8d$. This number of strategies on the CGLMP facet is larger than D for $d > 2$. For the case $d = 3$, the list is exhaustive, but not in general.

The second fact is that no extremal deterministic strategy belongs to the slice (5.53), as the marginals in the slice are completely random. Since it's a requirement of the present work that the final distribution belongs to the slice, Eve must manage to send deterministic strategies with the suitable probabilities in order to pertain to the slice. Therefore, at least one local point on the slice can be obtained by several different decompositions on extremal points and the decomposition chosen must be the one that optimizes Eve's information.

As a third fact, we analyze the deterministic strategies that are more interesting for Eve, in fact, two kind of local points are of special interest for her:

(i) those for which $b(0) = b(1)$, because Eve knows Bob's symbol when Alice announces $x = 0$, and which we denote by the set $\mathcal{L}_0$. In this case, looking at (5.61) the last two conditions become equal and the δ's compensate each other for all $a(1)$, so the only way to saturate the inequality is to fulfill both $b(0) \geq a(0)$ and $a(0) \geq b(1) = b(0)$; whence $a(0) = b(0) = b(1)$. These points are d^2 in total: those for which $a(0) = b(0) = b(1)$ and $a(1)$ can take any value. In other words, there are no points on the CGLMP facet such that $b(0) = b(1)$ but $a(0)$ is different from this value: whenever Eve learns Bob's symbol for $x = 0$, Alice and Bob make no error for $x = 0$.

(ii) those for which $b(0) = b(1) - 1$, because Eve knows Bob's symbol when Alice announces $x = 1$, and which we denote by the set $\mathcal{L}_1$. If one first considers $b(0) < d - 1$ then $b(1) > b(0)$, therefore the first two conditions cannot be simultaneously fulfilled, whatever $a(0)$ is. One can then verify that only the choice $a(1) = b(0)$ leads to a saturation of the inequality, as the last term of (5.61) is not fulfilled. The last remaining case is $b(0) = d - 1$, $b(1) = 0$: it can be read directly from the array, and leads to the same conclusion. There are also d^2 points in $\mathcal{L}_1$: those for which $a(1) = b(0) = b(1) - 1$ and $a(0)$ can take any value. This has a similar interpretation as the statement above, in the case $x = 1$.

In all the other cases that are not in $\mathcal{L} \equiv \mathcal{L}_0 \cup \mathcal{L}_1$, Eve does not learn Bob's symbol with certainty. Now, since the error rate Alice-Bob depends only on $P(a, b|x, y)$, and not on the particular decomposition chosen by Eve to realize this distribution, it is obvious that Eve's best strategy lies in distributing local points that belong to $\mathcal{L} \equiv \mathcal{L}_0 \cup \mathcal{L}_1$ as often as possible. For $d = 3$, we shall prove that she can prepare any point in the slice by distributing only these kind of local points. An additional distinction within $\mathcal{L}$ will be explicitly detailed in the case $d = 3$. We shall call $\mathcal{L}^3$ the subset of $\mathcal{L}$, whose points satisfy three out of the four relations $a(0) = b(0)$, $a(0) = b(1)$, $a(1) = b(0)$ and $a(1) = b(1) - 1$; the complementary set, containing the points that satisfy two of the relations ($a(0) = b(0) = b(1)$ or $a(1) = b(0) = b(1) - 1$), is written $\mathcal{L}^2$.

Eve's strategy: non-local point

When we described the non-signaling polytope for d outcomes in chapter 4 we said that among the extremal non-local points which lie above the CGLMP facet, the only one on the slice (5.53) is $PR_{2,d}$. However, it may be the case that mixtures of other extremal non-local points lie as well in the slice. For $d = 3$, this is not the case.
In this study, we suppose that Eve sends a unique non-local strategy, namely $PR_{2,d}$. Under this assumption, we can define p_{NL} as the probability that Eve sends $PR_{2,d}$. To find the expression of p_{NL}, we notice that $\langle I_d, PR_{2,d}\rangle = \frac{d-1}{d}$ should correspond to $p_{NL} = 1$, and that $\langle I_d, L\rangle = 0$ for all local points on the CGLMP facet, should correspond to $p_{NL} = 0$. Moreover, p_{NL} measures the geometrical distance from the facet and is therefore an affine function of the violation of CGLMP. Thus for a generic distribution P of the form (4.36) we have

$$\begin{aligned} p_{NL} &= \frac{d}{d-1}\,\langle I_d, P\rangle = \\ &= -2 + \sum_{\Delta=0}^{d-1}\left(1 - \frac{\Delta}{d-1}\right)\left[3p_{-\Delta} - p_{\Delta+1}\right]. \end{aligned} \tag{5.64}$$

Now we can present the results for the possibility of extracting a secret key from a detailed study of the case $d = 3$.

5.4.2 Secret key extraction: $d = 3$

The slice of the polytope

The slice (5.53) is 2-dimensional for $d = 3$, we choose p_0 and p_1 as free parameters; this gives $p_2 = 1 - p_0 - p_1$ and

$$p_{NL} = -2 + (3p_0 - p_1) + p_2 = 2(p_0 - p_1) - 1\,. \tag{5.65}$$

The full slice has a form of an equilateral triangle (Fig. 5.4), whose vertices V_Δ are defined by $p_\Delta = 1$. As mentioned, $V_0 = PR_{2,3}$. The vertex V_2 is also a $PR_{2,3}$, the one defined by $b - a = \bar{x}\bar{y} + 1$ with $\bar{z} = 1 - z$. On the contrary, V_1 a mixture of deterministic strategies. The middle of the triangle, $p_0 = p_1 = p_2 = \frac{1}{3}$, is the completely random strategy.

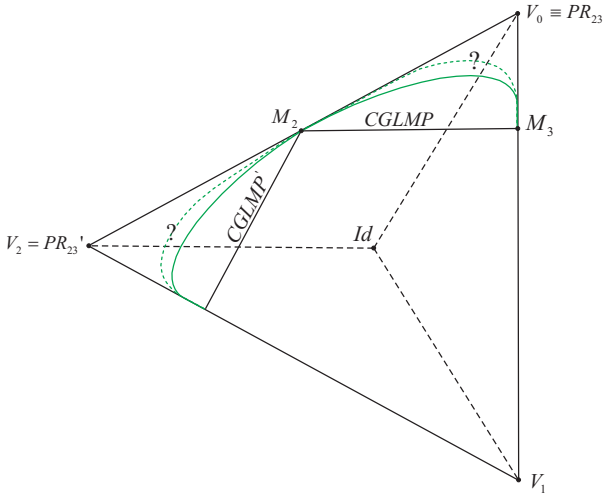


Figure 5.4: The slice (5.53) of the no-signalling polytope, for $d = 3$. The full extent of the quantum region is not known, it is represented by the dotted line with question marks. The full line is the part of the quantum region that can certainly reach and that was described in detail in [18]

Explicit analysis for $d = 3$

Deterministic strategies on the facet

We give here the explicit list of the 30 $(6d^2 - 8d)$ deterministic strategies that saturate CGLMP. As stated before, there are $4d = 12$ strategies that fulfill three of the relationships and thus $|\mathcal{L}^3| = 12$. As the size of $\mathcal{L}$ is $2d^2$, there are 6 strategies in $\mathcal{L}^2$. We note $r = 0, 1, 2$.

The twelve strategies in $\mathcal{L}^3$ are

$$\mathcal{L}_0^3 \;:\; \begin{cases} L_{3,1}^r &= \{a(x) = r, b(y) = r\} \\ L_{3,2}^r &= \{a(x) = r - x, b(y) = r\} \end{cases} \tag{5.66}$$

$$\mathcal{L}_1^3 \;:\; \begin{cases} L_{3,3}^r &= \{a(x) = r, b(y) = r + y\} \\ L_{3,4}^r &= \{a(x) = r - x, b(y) = r + y - 1\} \end{cases} . \tag{5.67}$$

The six strategies in $\mathcal{L}^2$ are

$$\mathcal{L}_0^2 \;:\; L_{2,1}^r = \{a(x) = r + x, b(y) = r\} \tag{5.68}$$

$$\mathcal{L}_1^2 \;:\; L_{2,2}^r = \{a(x) = r + x, b(y) = r + y + 1\} . \tag{5.69}$$

The twelve strategies outside $\mathcal{L}$ are:

$$\begin{array}{rcl} L_{e,1}^r &=& \{a(x) = r, b(y) = r - y\} \\ L_{e,2}^r &=& \{a(x) = r + x, b(y) = r - y\} \\ L_{e,3}^r &=& \{a(x) = r + x, b(y) = r - y + 1\} \\ L_{e,4}^r &=& \{a(x) = r - x, b(y) = r - y + 1\} . \end{array} \tag{5.70}$$

In Table 5.6 we give the probabilities of Alice's and Bob's symbols for each of the strategies that can be sent by Eve.

Onwards, we are going to focus on the non-local region close to V_0 (Fig. 5.5). The intersection with the CGLMP facet is the segment $p_0 - p_1 = \frac{1}{2}$, whose ends are the points labeled M_2 ($p_0 = \frac{1}{2}$, $p_1 = 0$) and M_3 ($p_0 = \frac{3}{4}$, $p_1 = \frac{1}{4}$). The decompositions chosen of these mixtures on the extremal deterministic strategies are

$$M_2 = \sum_{L \in \mathcal{L}^2} \frac{1}{6} L \quad , \quad M_3 = \sum_{L \in \mathcal{L}^3} \frac{1}{12} L \tag{5.71}$$

where the sets of local points $\mathcal{L}^2$ and $\mathcal{L}^3$ have been defined above.

In fact, the decomposition of M_3 is unique; conversely, M_2 can be decomposed in an infinity of ways, but all the others involve also the points that don't belong to $\mathcal{L}$ and are therefore sub-optimal for Eve.

The general decomposition of M_2 is defined by

$$M_2 \;:\; \begin{array}{rcl} p_{3,j}^r &=& 0 , \\ p_{2,1}^r &=& p_{2,2}^r \equiv p_2^r \text{ free}, \\ p_{e,1}^r &\text{free},& \\ p_{e,2}^r &=& \frac{1}{6} - (p_2^r + p_{e,1}^r) , \\ p_{e,3}^r &=& p_{e,1}^{r+1} , \\ p_{e,4}^r &=& \frac{1}{6} - (p_2^r + p_{e,1}^{r+1}) . \end{array} \tag{5.72}$$

There are thus six free parameters $\{p_2^r, p_{e,1}^r\}$, constrained of course by the positivity of probabilities (in particular, none of these parameters can exceed $\frac{1}{6}$). A possible realization of M_2 is the equiprobable mixture of the eighteen points which are not in $\mathcal{L}^3$. The choice leading to (5.71) is the equiprobable mixture of the six points in $\mathcal{L}^2$ that give Eve the highest advantage($p_2^r = \frac{1}{6}$, implying automatically $p_{e,j}^r = 0$).
The Table for the correlations Alice-Bob-Eve can be given in general as it is shown in Table 5.7, where the notations

$$f(p) = \frac{2p_1}{3} + 2p_2 p \quad , \quad g(p) = \frac{1 - p_0}{3} - 2p_2 p . \tag{5.73}$$

$A \setminus B$	$y=0$ b=0	$y=0$ b=1	$y=0$ b=2	$y=1$ b=0	$y=1$ b=1	$y=1$ b=2
$x=0$, $a=0$	$\frac{p_{NL}}{3}$ (PR) $p^0_{3,1}$ $(L^0_{3,1})$ $p^0_{3,2}$ $(L^0_{3,2})$ $p^0_{3,3}$ $(L^0_{3,3})$ $p^0_{2,1}$ $(L^0_{2,1})$ $p^0_{e,1}$ $(L^0_{e,1})$ $p^0_{e,2}$ $(L^0_{e,2})$	$p^0_{2,2}$ $(L^0_{2,2})$ $p^0_{e,3}$ $(L^0_{e,3})$ $p^0_{e,4}$ $(L^0_{e,4})$	$p^0_{3,4}$ $(L^0_{3,4})$	$\frac{p_{NL}}{3}$ (PR) $p^0_{3,1}$ $(L^0_{3,1})$ $p^0_{3,2}$ $(L^0_{3,2})$ $p^0_{3,4}$ $(L^0_{3,4})$ $p^0_{2,1}$ $(L^0_{2,1})$ $p^0_{e,3}$ $(L^0_{e,3})$ $p^0_{e,4}$ $(L^0_{e,4})$	$p^0_{3,3}$ $(L^0_{3,3})$	$p^0_{2,2}$ $(L^0_{2,2})$ $p^0_{e,1}$ $(L^0_{e,1})$ $p^0_{e,2}$ $(L^0_{e,2})$
$x=0$, $a=1$,	$p^1_{3,4}$ $(L^1_{3,4})$	$\frac{p_{NL}}{3}$ (PR) $p^1_{3,1}$ $(L^1_{3,1})$ $p^1_{3,2}$ $(L^1_{3,2})$ $p^1_{3,3}$ $(L^1_{3,3})$ $p^1_{2,1}$ $(L^1_{2,1})$ $p^1_{e,1}$ $(L^1_{e,1})$ $p^1_{e,2}$ $(L^1_{e,2})$	$p^1_{2,2}$ $(L^1_{2,2})$ $p^1_{e,3}$ $(L^1_{e,3})$ $p^1_{e,4}$ $(L^1_{e,4})$	$p^1_{2,2}$ $(L^1_{2,2})$ $p^1_{e,1}$ $(L^1_{e,1})$ $p^1_{e,2}$ $(L^1_{e,2})$	$\frac{p_{NL}}{3}$ (PR) $p^1_{3,1}$ $(L^1_{3,1})$ $p^1_{3,2}$ $(L^1_{3,2})$ $p^1_{3,4}$ $(L^1_{3,4})$ $p^1_{2,1}$ $(L^1_{2,1})$ $p^1_{e,3}$ $(L^1_{e,3})$ $p^1_{e,4}$ $(L^1_{e,4})$	$p^1_{3,3}$ $(L^1_{3,3})$
$x=0$, $a=2$,	$p^2_{2,2}$ $(L^2_{2,2})$ $p^2_{e,3}$ $(L^2_{e,3})$ $p^2_{e,4}$ $(L^2_{e,4})$	$p^2_{3,4}$ $(L^2_{3,4})$	$\frac{p_{NL}}{3}$ (PR) $p^2_{3,1}$ $(L^2_{3,1})$ $p^2_{3,2}$ $(L^2_{3,2})$ $p^2_{3,3}$ $(L^2_{3,3})$ $p^2_{2,1}$ $(L^2_{2,1})$ $p^2_{e,1}$ $(L^2_{e,1})$ $p^2_{e,2}$ $(L^2_{e,2})$	$p^2_{3,3}$ $(L^2_{3,3})$	$p^2_{2,2}$ $(L^2_{2,2})$ $p^2_{e,1}$ $(L^2_{e,1})$ $p^2_{e,2}$ $(L^2_{e,2})$	$\frac{p_{NL}}{3}$ (PR) $p^2_{3,1}$ $(L^2_{3,1})$ $p^2_{3,2}$ $(L^2_{3,2})$ $p^2_{3,4}$ $(L^2_{3,4})$ $p^2_{2,1}$ $(L^2_{2,1})$ $p^2_{e,3}$ $(L^2_{e,3})$ $p^2_{e,4}$ $(L^2_{e,4})$
$x=1$, $a=0$,	$\frac{p_{NL}}{3}$ (PR) $p^0_{3,1}$ $(L^0_{3,1})$ $p^0_{3,3}$ $(L^0_{3,3})$ $p^1_{3,4}$ $(L^1_{3,4})$ $p^2_{2,2}$ $(L^2_{2,2})$ $p^0_{e,1}$ $(L^0_{e,1})$ $p^2_{e,3}$ $(L^2_{e,3})$	$p^1_{3,2}$ $(L^1_{3,2})$	$p^2_{2,1}$ $(L^2_{2,1})$ $p^2_{e,2}$ $(L^2_{e,2})$ $p^1_{e,4}$ $(L^1_{e,4})$	$\frac{p_{NL}}{3}$ (PR) $p^1_{3,2}$ $(L^1_{3,2})$	$p^0_{2,1}$ $(L^0_{2,1})$ $p^0_{3,3}$ $(L^0_{3,3})$ $p^1_{3,4}$ $(L^1_{3,4})$ $p^2_{2,2}$ $(L^2_{2,2})$ $p^2_{e,2}$ $(L^2_{e,2})$ $p^1_{e,4}$ $(L^1_{e,4})$	$p^0_{3,1}$ $(L^0_{3,1})$ $p^0_{e,1}$ $(L^0_{e,1})$ $p^2_{e,3}$ $(L^2_{e,3})$
$x=1$, $a=1$,	$p^0_{2,1}$ $(L^0_{2,1})$ $p^0_{e,2}$ $(L^0_{e,2})$ $p^2_{e,4}$ $(L^2_{e,4})$	$\frac{p_{NL}}{3}$ (PR) $p^1_{3,1}$ $(L^1_{3,1})$ $p^1_{3,3}$ $(L^1_{3,3})$ $p^2_{3,4}$ $(L^2_{3,4})$ $p^0_{2,2}$ $(L^0_{2,2})$ $p^1_{e,1}$ $(L^1_{e,1})$ $p^0_{e,3}$ $(L^0_{e,3})$	$p^2_{3,2}$ $(L^2_{3,2})$ $p^2_{3,2}$ $(L^2_{3,2})$	$p^1_{3,1}$ $(L^1_{3,1})$ $p^1_{e,1}$ $(L^1_{e,1})$ $p^0_{e,3}$ $(L^0_{e,3})$	$\frac{p_{NL}}{3}$ (PR) $p^2_{3,2}$ $(L^2_{3,2})$	$p^0_{2,1}$ $(L^0_{2,1})$ $p^1_{3,3}$ $(L^1_{3,3})$ $p^2_{3,4}$ $(L^2_{3,4})$ $p^0_{2,2}$ $(L^0_{2,2})$ $p^0_{e,2}$ $(L^0_{e,2})$ $p^2_{e,4}$ $(L^2_{e,4})$
$x=1$, $a=2$,	$p^0_{3,2}$ $(L^0_{3,2})$	$p^1_{2,1}$ $(L^1_{2,1})$ $p^1_{e,2}$ $(L^1_{e,2})$ $p^0_{e,4}$ $(L^0_{e,4})$	$\frac{p_{NL}}{3}$ (PR) $p^2_{3,1}$ $(L^2_{3,1})$ $p^2_{3,3}$ $(L^2_{3,3})$ $p^0_{3,4}$ $(L^0_{3,4})$ $p^1_{2,2}$ $(L^1_{2,2})$ $p^2_{e,1}$ $(L^2_{e,1})$ $p^1_{e,3}$ $(L^1_{e,3})$	$p^1_{2,1}$ $(L^1_{2,1})$ $p^2_{3,3}$ $(L^2_{3,3})$ $p^0_{3,4}$ $(L^0_{3,4})$ $p^1_{2,2}$ $(L^1_{2,2})$ $p^1_{e,2}$ $(L^1_{e,2})$ $p^0_{e,4}$ $(L^0_{e,4})$	$p^2_{3,1}$ $(L^2_{3,1})$ $p^1_{e,1}$ $(L^1_{e,1})$ $p^1_{e,3}$ $(L^1_{e,3})$	$\frac{p_{NL}}{3}$ (PR) $p^0_{3,2}$ $(L^0_{3,2})$

Table 5.6: Table of the distribution Alice-Bob-Eve for the raw data. The entries are the $P(a,b|x,y)$ for $d=3$ In parentheses, we indicate Eve's symbol. I red we represent the effect of sifting in the case $x=y=1$ which implies $b \to b-1$

x	$b=0$	$b=1$	$b=2$
$a=0$	$\frac{p_{NL}}{3}$ $(?,?)$ $f(p_2^{0-x})$ $(0,0)$ $g(p_2^{0-x})$ $(0,?_2)$	$\frac{1-p_0}{6}$ $(0,?_2)$	$\frac{1-p_0}{6}$ $(0,?_2)$
$a=1$	$\frac{1-p_0}{6}$ $(1,?_2)$	$\frac{p_{NL}}{3}$ $(?,?)$ $f(p_2^{1-x})$ $(1,1)$ $g(p_2^{1-x})$ $(1,?_2)$	$\frac{1-p_0}{6}$ $(1,?_2)$
$a=2$	$\frac{1-p_0}{6}$ $(2,?_2)$	$\frac{1-p_0}{6}$ $(2,?_2)$	$\frac{p_{NL}}{3}$ $(?,?)$ $f(p_2^{2-x})$ $(2,2)$ $g(p_2^{2-x})$ $(2,?_2)$

Table 5.7: Probability distribution Alice-Bob-Eve for the sifted data, for $d=3$ an Alice's setting x, for the general decomposition (5.72) of M_2.

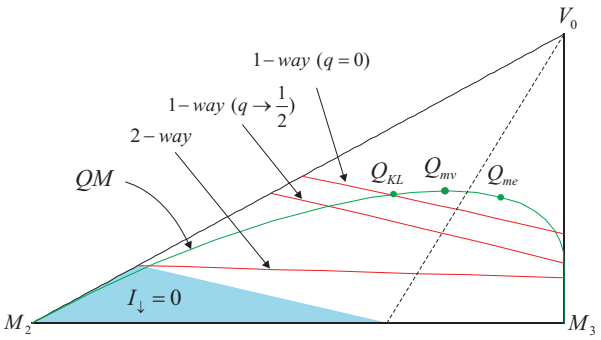


Figure 5.5: Zoom of Fig. 5.4 on the non-local region close to V_0. Only the part of the quantum region that we consider is represented here. The transverse lines define the limits down to which secrecy can be extracted for one-way post-processing (without and with pre-processing) and for two-way post-processing without pre-processing. In the shaded region, the intrinsic information $\mathrm{I}(A;B\downarrow E)$ is zero. We stress that this figure is an exact plot, not just an "artist view". See text for the other details.

are employed. This table is derived from Table 5.6 and the general decomposition of M_2. Note that, in each of the nine cells, the sum of the probabilities does not depend on the p_2^r, as $f(p)+g(p)$ does not depend on p. This is because the decomposition of M_2 is known only to Eve and Eve best strategy is maximizing the probability of knowing both symbols, measured by $f(p)$; whence the choice $p_2^r=\frac{1}{6}$ made. The symbol $?_2$ has been introduced to describe the situation where Eve is uncertain on Bob's symbol, but only among two possibilities: this is clearly the case whenever the uncertainty derives from a deterministic strategy.

The quantum-mechanical studies (see Appendix B in [18] for more details) have singled out two non-local probability distributions in this region. Their depiction is shown as an illustration of quantum mechanical states in the slice of the polytope. The first one, Q_{mv}, corresponds to the maximal violation of CGLMP using two qutrits, $p_{NL}\approx 0.4574$. The second one, Q_{me} corresponds to the highest violation achievable with the maximally entangled state of two qutrits, $p_{NL}\approx 0.4365$.

One-way classical post-processing

The table for the correlations Alice-Bob-Eve with the specified decomposition of M_2 showing the corresponding information that Eve can extract is Table 5.8. This table is similar to the one given in the case $d=2$ but more assumptions and analysis of the structure of the polytope has been needed.

All the probabilities in a row/column sum up to $\frac{1}{3}$ and we have the following values for the error between Alice and Bob.

$$e_{AB}(+1)\ =\ e_{AB}(-1)\quad =\quad \frac{1-p_0}{2} \tag{5.74}$$

	$b=0$	$b=1$	$b=2$
$a=0$	$\frac{p_{NL}}{3}\ (?,?)$ $\frac{p_L}{6}\ (0,0)$ $\frac{p_L}{12}-\frac{p_2}{6}\ (0,?_2)$	$\frac{1-p_0}{6}\ (0,?_2)$	$\frac{1-p_0}{6}\ (0,?_2)$
$a=1$	$\frac{1-p_0}{6}\ (1,?_2)$	$\frac{p_{NL}}{3}\ (?,?)$ $\frac{p_L}{6}\ (1,1)$ $\frac{p_L}{12}-\frac{p_2}{6}\ (1,?_2)$	$\frac{1-p_0}{6}\ (1,?_2)$
$a=2$	$\frac{1-p_0}{6}\ (2,?_2)$	$\frac{1-p_0}{6}\ (2,?_2)$	$\frac{p_{NL}}{3}\ (?,?)$ $\frac{p_L}{6}\ (2,2)$ $\frac{p_L}{12}-\frac{p_2}{6}\ (2,?_2)$

Table 5.8: Probability distribution Alice-Bob-Eve for the sifted data, for $d=3$, assuming decomposition (5.71) for M_2.

as expected from (5.60). In all that follows, information is quantified in trits, and we write the ternary entropy as $h\big([v_1,v_2,v_3]\big)=-\sum_j v_j\log_3 v_j$.

Therefore we can obtain the estimate for the CK bound can be obtained. We make the calculation explicitly. We recall that in this case $r_{CK}=\mathrm{I}(B;A)-\mathrm{I}(B;E)$. In the case of $\mathrm{I}(A;B)$ employing $\mathrm{I}(A;B)=\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}}P(a,b)\log_3\dfrac{P(a,b)}{P_A(a)P_B(b)}$ we obtain

$$\mathrm{I}(A;B)=p_0\log_3(3p_0)+(1-p_0)\log_3\left(3\frac{1-p_0}{2}\right)$$

. To obtain $\mathrm{I}(B;E)$, reviewing the table of correlations, one has to take into account that in the absence of pre-processing we have the following:

- Eve in the cases where $(b,e)=(x,(?,?))$ has no information and this happens probability p_{NL}
- Eve has full information in the cases where $(b,e)\in\{(0,(0,0)),(1,(1,1)),(2,(2,2))\}$ which happen with probability $\frac{p_L}{2}$
- In the rest of cases, she is uncertain about Bob's symbol between two possible values. Therefore, when we compute for example the term in mutual information of $(b,e)=((0,(0,?_2))$, we have to consider just one of the possible values and both in $P(e)$, therefore the term is $(\frac{p_L}{12}-\frac{p_2}{6})\log_3\left(\frac{\frac{p_L}{12}-\frac{p_2}{6}}{\frac{1}{3}2\cdot(\frac{p_L}{12}-\frac{p_2}{6})}\right)$ and therefore, as $\frac{p_L}{12}-\frac{p_2}{6}=\frac{p_1}{3}$, the term becomes $\frac{p_1}{3}\log_3(\frac{3}{2})$. There are 3 of these terms.
  In the cases of symbols of the form $(b,e)=(x,(y,?_2)$ as for example $(b,e)=(0,(1,?_2))$, the situation is analog: the term in the mutual information is $\frac{1-p_0}{6}\log_3(\frac{3}{2})$. There are 6 of these terms. Adding the different terms, one has that the mutual information of theses cases is

$$p_1\log_3(\frac{3}{2})+(1-p_0)\log_3(\frac{3}{2})=\frac{p_L}{2}(1-\log_3 2)$$

  That is, she has information $1-h\big([1/2,1/2,0]\big)=1-\log_3 2$ with probability $\frac{p_L}{2}$.

Therefore

$$\begin{aligned}R_{CK}(q=0)\ \geq\ r_{CK}\ &=\ 1+\log_3 p_0+(1-p_0)\log_3(\frac{1-p_0}{2})-\frac{p_L}{2}-\frac{p_L}{2}(1-\log_3 2)\\ &=\ 1-h\left([p_0,\frac{1-p_0}{2},\frac{1-p_0}{2}]\right)-p_L\left(1-\frac{1}{2}\log_3 2\right).\end{aligned}\tag{5.75}$$

The curve $r_{CK}(q=0)=0$ is shown in Fig. 5.5, it clearly cuts the quantum region.

A natural question is, which is the point that maximizes $r_{CK}(q=0)$ under the requirement that the correlations should belong to the quantum region. That is, the intersection between the known quantum region and the line given by $r_{CK}(q=0)$. In the slice under consideration, the rate was

found in [18] $r_{max}(q=0) \approx 0.09$ trits ≈ 0.144 bits for the correlations defined by $p_0 = 0.8286$, $p_1 = 0.1093$. These correlations can be obtained by measuring the quantum state

$$|\psi(\gamma)\rangle \quad = \quad \frac{1}{\sqrt{2+\gamma^2}}\left(|00\rangle + \gamma|11\rangle + |22\rangle\right) \tag{5.76}$$

for $\gamma \approx 0.9875$. This state is close to, but different from, the maximally entangled state.

Considering Bob's pre-processing, we must first consider that for one-way post-processing, dit-wise pre-processing is already optimal. One can define two different flipping probabilities q_{+1} and q_{-1}, associated respectively to $b \to b+1$ and $b \to b+2$. But inspecting the correlations and taking into account their symmetries, the optimal is always obtained for $q_{+1} = q_{-1} = q$. Taking into account this, we have that the error between Alice and Bob becomes:

$$e'_{AB}(+1) = e'_{AB}(-1) \quad \equiv \quad \frac{e'}{2} = 3\left((1-q)\frac{1-p_0}{6} + q\frac{p_0}{3} + \frac{1-p_0}{6}\right) = \frac{1-p_0}{2} + q\,\frac{3p_0-1}{2}$$

where we have used that $\frac{p_{NL}}{3} + \frac{p_L}{6} + \frac{p_L}{12} - \frac{p_2}{6} = \frac{p_0}{3}$
And therefore, mutual information can be obtained as:

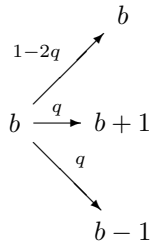
$$I(A:B') \quad = \quad 1 - h\left([1-e', \frac{e'}{2}, \frac{e'}{2}]\right) . \tag{5.77}$$

Eve's information is computed by recalling the analysis made before

- In the case she sends the non-local point, she has no information.
- In case she sends out a local point, for one value of x, she knows perfectly Bob's symbol b. Pre-processing leaves b unchanged with probability $1-2q$, and sends it to $b\pm1$ with probability q each. The component of mutual information in this case is

  $$(1-2q)\frac{p_L}{2}\log_3(3(1-2q)) + 2q\frac{p_L}{2}\log_3(3q) \tag{5.78}$$

  Thus Eve's information is lowered from 1 to $1 - h([1-2q, q, q])$.
- For the other value of x, she is uncertain between two values of b. Considering Bob's flips:



  Eve's random guess in $(x, ?_2)$ will be right with probability $\frac{1-2q}{2} + \frac{q}{2} = \frac{1-q}{2}$. Eve's information is lowered from the previous $1 - h([\frac{1}{2}, \frac{1}{2}, 0])$ to $1 - h([\frac{1-q}{2}, \frac{1-q}{2}, q])$.

Since each case of the local points analyzed is equiprobable,

$$\mathrm{I}(E;B') \quad = \quad p_L\left(1 - \frac{1}{2}\left[h([1-2q,q,q]) + h([\frac{1-q}{2}, \frac{1-q}{2}, q])\right]\right) . \tag{5.79}$$

From (5.77) and (5.79), r_{CK} can be computed by optimizing the value of q. Doing the optimization numerically, in Figure 5.5 one can see that the improvement due to pre-processing is clear.

Two-way classical post-processing

The possibility of extracting a secret key from the correlations of Table 5.8 using advantage distillation without considering pre-processing has been studied. Alice selects N of her symbols that are

identical, Bob accepts if and only if his corresponding symbols are also identical. The probability that Bob accepts is $p_0^N + [e_{AB}(+1)]^N + [e_{AB}(-1)]^N$, and consequently

$$\tilde{e}_{AB}(\pm 1) \quad = \quad \frac{\left(\frac{1-p_0}{2}\right)^N}{p_0^N + 2\left(\frac{1-p_0}{2}\right)^N} = \frac{(1-p_0)^N}{2^N p_0^N + 2(1-p_0)^N} \approx \left(\frac{1-p_0}{2p_0}\right)^N . \tag{5.80}$$

As in the case $d = 2$, Eve has to make a random guess if and only if she has sent $PR_{2,3}$ for all the N instances:

$$\tilde{e}_E(\pm 1) \quad \gtrsim \quad \frac{1}{3}\left(\frac{p_{NL}}{p_0}\right)^N . \tag{5.81}$$

Therefor, a secret key can be extracted using advantage distillation as long as $p_{NL} > \frac{1-p_0}{2}$, that is as long as

$$5p_0 \quad > \quad 4p_1 + 3\,. \tag{5.82}$$

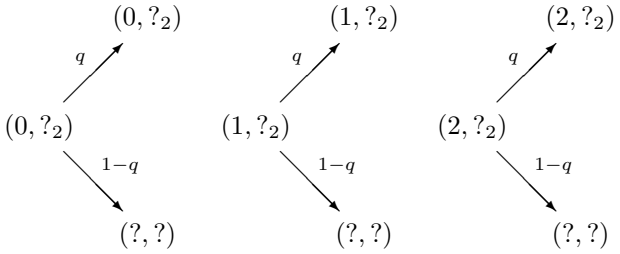
The limiting curve is plotted in Figure 5.5. Its extremal points are $p_{NL} > \frac{1}{5}$ for $p_1 = 0$ and $p_{NL} > \frac{1}{9}$ for $p_2 = 0$.

Intrinsic information $d = 3$

From Table 5.8 we can extract that the mutual conditioned information in this case is:

$$\mathrm{I}(A;B \mid E) = (2(p_0 - p_1) - 1) = p_{NL} \tag{5.83}$$

This is exactly what happened in the case of $d = 2$. In this case, the numerical optimization is harder because there are more symbols to consider and because their probabilities depend of 2 parameters. Intuitively, one can think that the maps of the form:



will also give us the intrinsic information as a generalization of the case of $d = 2$.

	$b=0$	$b=1$	$b=2$
$a=0$	$\frac{2(p_0-p_1)-1}{3} + (1-q)\frac{p_1}{3}$ $(?,?)$ $\frac{1+p_1-p_0}{3}$ $(0,0)$ $q\frac{p_1}{3}$ $(0,?_2)$	$(1-q)\frac{1-p_0}{6}$ $(?,?)$ $q\frac{1-p_0}{6}$ $(0,?_2)$	$(1-q)\frac{1-p_0}{6}$ $(?,?)$ $q\frac{1-p_0}{6}$ $(0,?_2)$
$a=1$	$(1-q)\frac{1-p_0}{6}$ $(?,?)$ $q\frac{1-p_0}{6}$ $(1,?_2)$	$\frac{2(p_0-p_1)-1}{3} + (1-q)\frac{p_1}{3}$ $(?,?)$ $\frac{1+p_1-p_0}{3}$ $(1,1)$ $q\frac{p_1}{3}$ $(1,?_2)$	$(1-q)\frac{1-p_0}{6}$ $(?,?)$ $q\frac{1-p_0}{6}$ $(1,?_2)$
$a=2$	$(1-q)\frac{1-p_0}{6}$ $(?,?)$ $q\frac{1-p_0}{6}$ $(2,?_2)$	$(1-q)\frac{1-p_0}{6}$ $(?,?)$ $q\frac{1-p_0}{6}$ $(2,?_2)$	$\frac{2(p_0-p_1)-1}{3} + (1-q)\frac{p_1}{3}$ $(?,?)$ $\frac{1+p_1-p_0}{6}$ $(2,2)$ $q\frac{p_1}{3}$ $(2,?_2)$

Table 5.9: Probability distribution Alice-Bob-Eve for the sifted data, for $d = 3$, assuming decomposition (5.71) for M_2, and the channel for Eve presented above. We express all joint probabilities as functions of p_0 and p_1

With these transformations we have that the mutual conditioned information in this case is:

$$\begin{aligned} \mathrm{I}(A;B\mid E)_q = \quad & ((2(p_0-p_1)-1)+(1-q)p_1+(1-q)(1-p_0)) \\ & (1-h[\tfrac{(2(p_0-p_1)-1)+(1-q)p_1}{((2(p_0-p_1)-1)+(1-q)p_1+(1-q)(1-p_0))}, \\ & \tfrac{\frac{1}{2}(1-q)p_0}{((2(p_0-p_1)-1)+(1-q)p_1+(1-q)(1-p_0))}, \\ & \tfrac{\frac{1}{2}(1-q)p_0}{((2(p_0-p_1)-1)+(1-q)p_1+(1-q)(1-p_0))}]) \end{aligned} \tag{5.84}$$

This case is different from the one of $d=2$, because with this expression one can find lines where $\mathrm{I}(A;B\mid E)_q=0$ of the form $p_1=\frac{1}{1+q}(\frac{5}{2}p_0-\frac{1}{2}qp_0+\frac{q}{2}-\frac{3}{2})$. This means that a positive key rate is impossible in the presence of these zeros.

Mapping all the symbols $(i,?_2)$ into $(?,?)$, where $i=0,1,2$, that is $q=0$. The obtained conditional mutual information reads

$$\begin{aligned} I(A:B|\bar{E}) \quad = \quad & P(?,?)\times \\ & \left[1-h\left(\frac{2p_0-p_L}{2-p_L},\frac{1-p_0}{2-p_L},\frac{1-p_0}{2-p_L}\right)\right]. \end{aligned} \tag{5.85}$$

Contrary to what happened in the $d=2$ case, this quantity vanishes for some points inside the region of Bell violation Indeed, Eq. (5.85) is zero on the line $5p_0-2p_1-3=0$. For different values of q in the map described, the intrinsic information is zero also below the line, that is for

$$5p_0-2p_1-3 \quad \leq \quad 0\,. \tag{5.86}$$

This region overlaps with the non-local region (Figure 5.6).

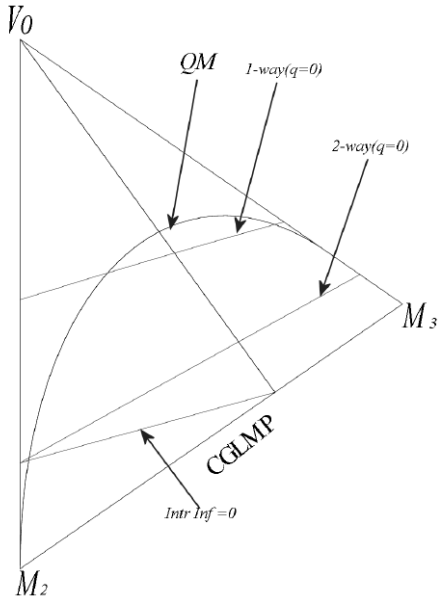


Figure 5.6: Intrinsic information $d=3$

5.4.3 Secret key extraction: generic d

Additional results were proven for a generic d in [18]. In summary, it was there conjectured that secrecy can be extracted from quantum non-local correlations for any d; and more precisely, that the amount of extractable secrecy increases when increasing d.

Chapter 6
Conclusions

The research in quantum information theory is intense and advancing in the two-fold goal of understanding reality and developing applications of quantum mechanics to information theory. Particularly, a lot has been done in the area of device independent cryptographic protocols based in quantum phenomena. The security in these protocols is based in the statistics observed by the honest parties when they interact with the devices' results, which allow them to decide if secrecy can be extracted.

A device providing secret key generation from the laws of quantum physics is for the parties a black box. The interactions of the parties with the device, the protocol, must be reliable and strong in order to guarantee security based exclusively in these interactions.

It's the phenomena of quantum non-locality that makes possible this feature: when a Bell inequality is violated with enough margin, secret shared randomness can be produced and therefore, security is guaranteed.

From the point of view of the project, the work done in 2006 was a first approach to the materialization of these concepts. The study of the CHSH protocol and its extension was made with restrictions in the assumptions: assuming individual attacks, bound to extract a secret key was obtained from some no-signaling probability distributions if a Bell inequality was violated. In particular, noisy quantum states can be used to distribute correlations, that are non-local enough to contain distillable secrecy.

The author could study concepts from classical and quantum information theory related to secrecy generation and interact with cutting-edge research at the moment.

6.1 Perspectives '06

The article published in 2006 ([18]) raised some perspectives that are worth to be cited here:

- Extend the security proofs to more general attacks by an eavesdropper limited by non-signaling.
- Develop device-independent proof of security against an eavesdropper which would be limited by quantum physics. From Alice and Bob perspective, non-locality should be the basis of security. On the side of Eve, the requirement that she must respect quantum physics is a limitation, compared to power we gave her in this paper; so one can hope to obtain a device-independent proof with better bounds.
- Look for optimal protocol in terms secret key extraction. The fact that all non-local probability distributions can generate secrecy implies that better procedures could be found to the ones described or that non-local distributions close to the local limit provide examples of bipartite bound information [35, 36].

6.2 Device independent quantum cryptography '16

The idea of developing quantum key distribution protocols that were independent of the device employed and were based in the violation of Bell inequalities was continued and perfected. The

following recall of the research that has been done until this year is extracted from [44]. A long line of research since the publication of [18] led to protocols and proof techniques that establishes non-vanishing key rates with noise tolerance in an i.i.d. setting (devices with no memory or no time dependent behavior). The adversaries supposed supra quantum in the work here presented, were analyzed assuming the limitations of quantum physics and extending the type of attacks from individual to general. In [43], the security of a DIQKD protocol based in the CHSH inequality was proven in the asymptotic limit, when the device is used $n \to \infty$ and under the i.i.d. assumption.

The maximum amount of randomness that can be generated from one system violating a specific Bell inequality by a given amount has been well studied.

The study of security proofs in each protocol are complex and further development is made in developing security proofs for the iid and non iid scenarios.

Appendix A

Codis utilitzats

A.1 Intrinsic information numerical analysis

Es presenta a continuació el codi emprat per tal de realitzar la optimització de la informació intrínseca en el cas $d = 2$.

A.2 Codi Matlab

Code intrzero.m: Optimization function of intrinsic information

```
function [x,fval] = intrzero(v)
global p
B=v
 Aeq=[1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0;
      0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 ;
      0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 ;
      0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 ;
      0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1  ];
b=[1; 1; 1; 1;1];
%lb=-ub;
%x0 = [1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/%5
    ↪ 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5];     % Make a starting guess
    ↪  at the solution

%x0 = [1 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 1 0.2 0.2 0.2 0.2
    ↪ 0.2 1 0.2 0.2 0.2 0.2 0.2 1];
x0=[1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1];
%x0=[1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1];
%x0 = [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1];
%x0=[1 0.3 0.2 0 0 0 0.7 0.2 0 0 0 0 0.2 0 0 0 0 0.2 0.7 0 0 0 0.2 0.3
    ↪ 1];
%x0 = [0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
    ↪ 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5];
options = optimset('LargeScale','off','DiffMaxChange',1e-6,'MaxSQPIter'
    ↪ ,15);

options = optimset('Display','iter','Diagnostics','on','TolX',1.0e-2);
%lb = [1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5
    ↪ 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5];%[0 0 0 0 0 0 0 0 0 0 0 0 0
    ↪ 0 0 0 0 0 0 0 0 0 0 0 0];
%lb =[0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001
    ↪ 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001
    ↪ 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001 0.0001];
```

```
%lb=[0.01 0.01 0.01 0.01 0.001 0.001 0.001 0.001 0.01 0.01 0.01 0.01
    ↪ 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01
    ↪ 0.01];
lb = [0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001
    ↪ 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001 0.001
    ↪  0.001 0.001 0.001];
%lb=[0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005
    ↪ 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005
    ↪ 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005 0.0005];
ub = [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1];
[x,fval] = fmincon(@objfun,x0,[],[],Aeq,b,lb,ub,[],options);

function f = objfun(x)

global p
% Nonlinear objective function
    % Variable objval shared with objfun and runsharedvalues

f=0;
for i=0:4
  f=f-(entr(x(1+5*i)*p/4+x(2+5*i)*p/4+x(3+5*i)*(1-p)/2)+entr(x(4+5*i)*p
      ↪ /4+x(5+5*i)*p/4+x(3+5*i)*(1-p)/2)+entr(x(1+5*i)*p/4+x(2+5*i)*p
      ↪ /8+x(4+5*i)*p/8+x(3+5*i)*(1-p)/2)+entr(x(5+5*i)*p/4+x(2+5*i)*p
      ↪ /8+x(4+5*i)*p/8+x(3+5*i)*(1-p)/2)-entr(x(1+5*i)*p/4+x(2+5*i)*p
      ↪ /8+x(3+5*i)*(1-p)/2)-entr(x(4+5*i)*p/8+x(5+5*i)*p/4+x(3+5*i)
      ↪ *(1-p)/2)-entr(p/8*x(2+5*i))-entr(p/8*x(4+5*i))-entr(x(1+5*i)*p
      ↪ /4+x(2+5*i)*p/4+x(3+5*i)*(1-p)+x(4+5*i)*p/4+x(5+5*i)*p/4));
  %repasar la definicion de inf mutua condicionada
end


    end



  end
```

Codi programa.m:

```
i=1;
sol=[];
pl=[];
calc=[];
opt=[];
global p
for p=0.5:0.01:1
[x,fval]=intrzero(p);
x;
pl(i)=p;
calc(i)=-(1/2+1/2*p)*log2((1/2+1/2*p))-(1-(1/2+1/2*p))*log2(1-(1/2+1/2*p
    ↪ ))
-1/4*(2-p)*(-p/(2-p)*log2(p/(2-p))-(1-p/(2-p))*log2(1-p/(2-p)));
sol(i)=fval;
opt(i)=(1-p/2)*(1+entr((1/2-3*p/8)/(1/2-p/4))+entr(p/8/(1/2-p/4)));
i=i+1;
end
sol
plot(pl,sol,'-b',pl,calc,':g',pl,opt,'-r')
```

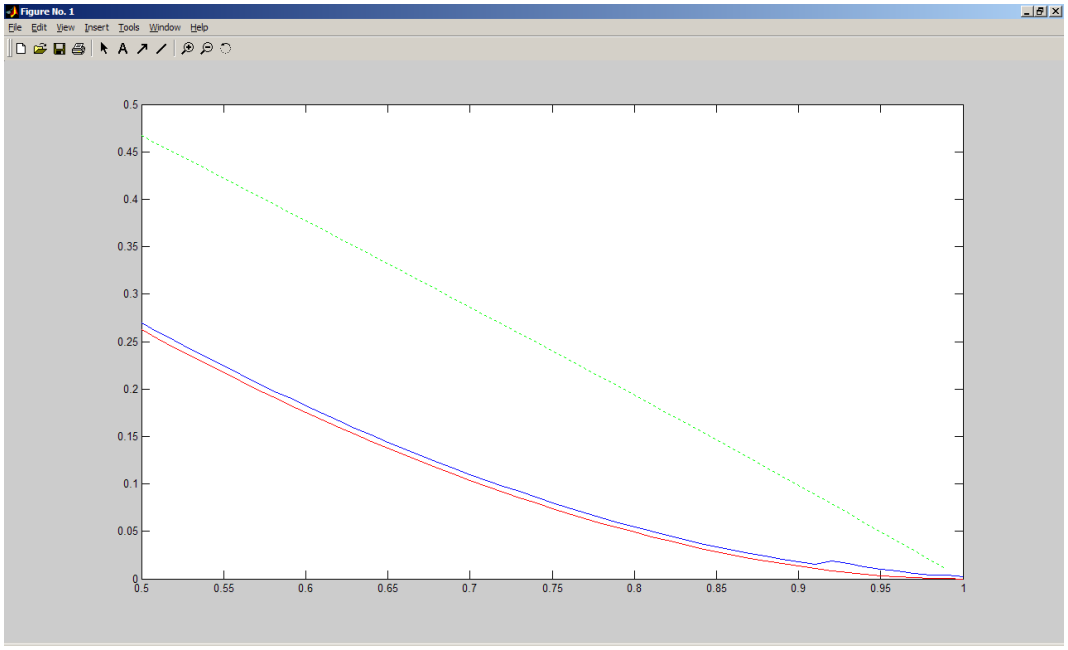


Figure A.1: Representation of the output of an example numerical optimization of the intrinsic information. No candidates were found with lower values thant the candidate proposed for $d = 2$

Codi calcul.m

```
global p
j=1;
sol=[];
pl=[];
gis=[];
zero=[];
%for p=0.999999:0.00000001:0.9999999999

%for y=0:0.1:1
%for z=0:0.1:1
%for t=0:0.1:1-z
%for p=0.999999:0.00000001:0.9999999999
for p=0.000001:0.001:1
%x=[1 0.7 0.2 0 0 0 0.3 0.2 0 0 0 0 0.2 0 0 0 0 0.2 0.3 0 0 0 0.2 0.7
    ↪ 1];
%x=[1 0 0 0 0 0 y 0 0 0 0 1-y 1 1-y 0 0 0 0 y 0 0 0 0 0 1];
%x = [1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5
    ↪ 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5 1/5];      % Make a starting guess
    ↪  at the solution
%x=[1 0.4 0.2 0 0 0 0.6 0.2 0 0 0 0 0.2 0 0 0 0 0.2 0.6 0 0 0 0.2 0.4
    ↪ 1];
%x=[1 0.3 0.2 0 0 0 0.7 0.2 0 0 0 0 0.2 0 0 0 0 0.2 0.7 0 0 0 0.2 0.3
    ↪ 1];
%x=[1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1];
%x=[1 0.5 0 0 0 0 0.5 0 0 0 0 0 1 0 0 0 0 0 0.5 0 0 0 0 0.5 1];
%x=[1 0 0 0 0 0 y 0 0 0 0 1-y 1 1-y 0 0 0 0 y 0 0 0 0 0 1];
%x=[(1-y)*(1-z) 0 0 0 0 y*(1-z) 1 0 0 0 z 0 1 0 z 0 0 0 1 y*(1-z) 0 0 0
    ↪ 0 (1-y)*(1-z)];
%x=[1 0 z/2 0 0 0 y z/2 0 0 0 1-y 1-z-t 1-y 0 0 0 t/2 y 0 0 0 t/2 0 1];
x=[1 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 1];
pl(j)=p;
```

```
f=0;
for  i=0:4
   f=f-(entr(x(1+5*i)*p/4+x(2+5*i)*p/4+x(3+5*i)*(1-p)/2)+entr(x(4+5*i)*p
       ↪ /4+x(5+5*i)*p/4+x(3+5*i)*(1-p)/2)+entr(x(1+5*i)*p/4+x(2+5*i)*p
       ↪ /8+x(4+5*i)*p/8+x(3+5*i)*(1-p)/2)+entr(x(5+5*i)*p/4+x(2+5*i)*p
       ↪ /8+x(4+5*i)*p/8+x(3+5*i)*(1-p)/2)-entr(x(1+5*i)*p/4+x(2+5*i)*p
       ↪ /8+x(3+5*i)*(1-p)/2)-entr(x(4+5*i)*p/8+x(5+5*i)*p/4+x(3+5*i)
       ↪ *(1-p)/2)-entr(p/8*x(2+5*i))-entr(p/8*x(4+5*i))-entr(x(1+5*i)*p
       ↪ /4+x(2+5*i)*p/4+x(3+5*i)*(1-p)+x(4+5*i)*p/4+x(5+5*i)*p/4));


end

gis(j)=-(1/2+1/2*p)*log2((1/2+1/2*p))-(1-(1/2+1/2*p))*log2(1-(1/2+1/2*p)
    ↪ )-1/4*(2-p)*(-p/(2-p)*log2(p/(2-p))-
(1-p/(2-p))*log2(1-p/(2-p)));
sol(j)=f;
zero(j)=(1-p/2)*(1+entr((1/2-3*p/8)/(1/2-p/4))+entr(p/8/(1/2-p/4)));
%zero(j)=(1-p/2*(1+y))*(1+entr((1/2-(3+y)*p/8)/(1/2-(1+y)*p/4))+entr((1-
    ↪ y)*p/8/(1/2-(1+y)*p/4)));
if  sol(j)<zero(j)
     print('Es menor');
%      y
 %     z
  %    t
     sol(j)
     zero(j)
end
j=j+1;



end
plot(pl,sol,'-r',pl,zero,'-b',pl,1-pl,':')%,pl,gis,':b'
end
end
end

%plot(pl,sol,'-r',pl,zero,pl,gis,':b')
%plot(pl,sol,pl,1-pl)
```

A.3 Maple codes

Representació de les fites d'informació per $d = 3$

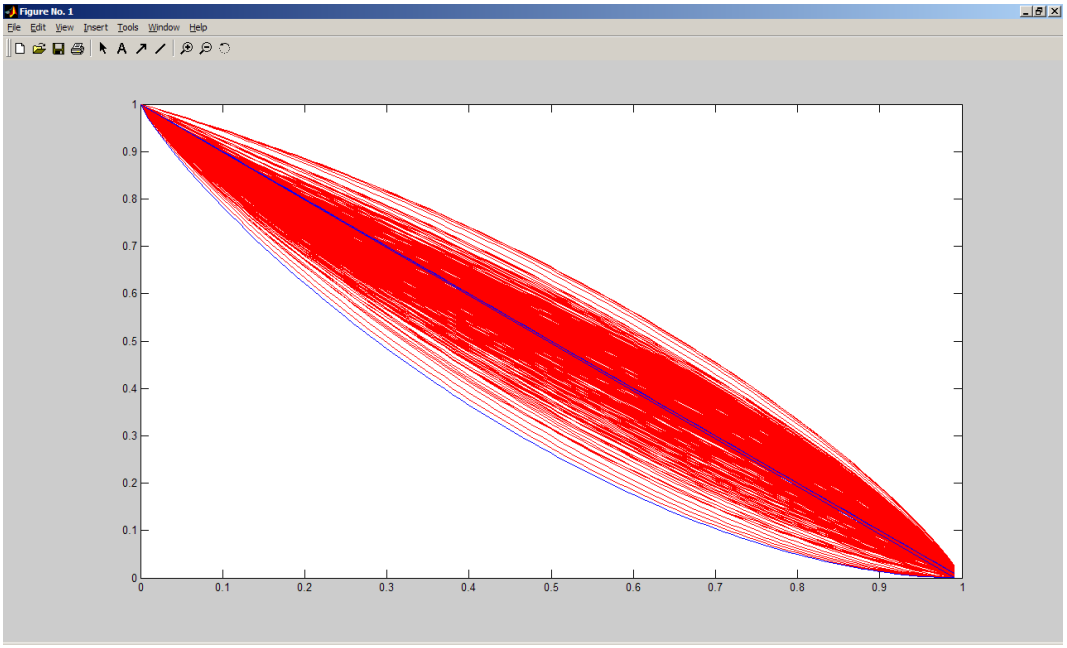


Figure A.2: Representation of the output of an example numerical optimization of the intrinsic information. No candidates were found with lower values thant the candidate proposed for $d = 2$

```
> A:=plot(x-0.5,x=0.5..0.75,color=blue):B:=plot(0,x=0.5..1,color=blue):C:=plot(1-x,x=0.75..1,color=blue)
  :E:=plot(g(x,0),x=0.5..2/3):U:=plot(w(x),x=0.5..0.8274412655):V:=plot(5/4*x-3/4,x=0.5..7/9):Q:=plot([1
  /3*(1+(1+2*sqrt(3)*t)/(2+t^2)),1/3*(1-(2)/(2+t^2)),
  t=0..1.8],color=black):W:=plot((1-x)/2,x=2/3..1,style=patch,color=black):
> display({A,B,C,E,U,V,Q,W},view=[0.5..1,0..0.25],axes=none);
>
```

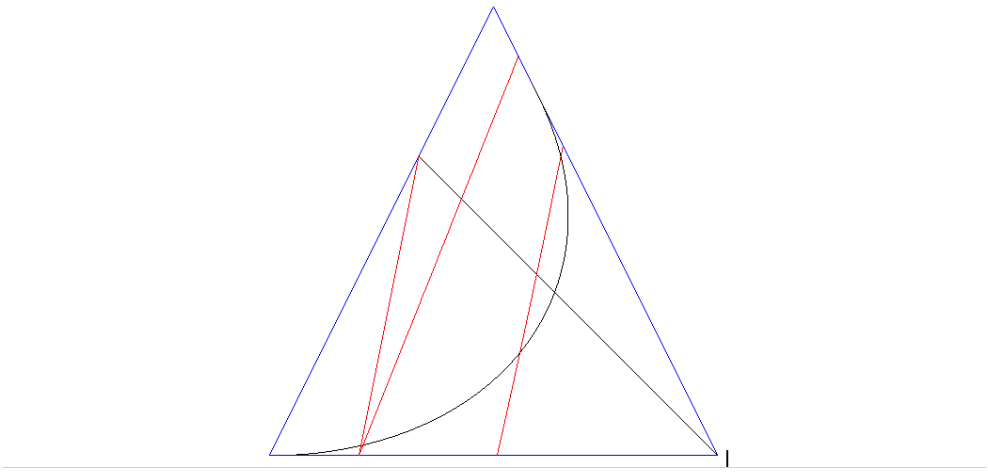


Figure A.3: Representation of secrecy extraction bounds and intrinsic information=0 in the slice of the polytope $d = 3$

Bibliography

[1] Peres, A., Quantum Theory: Concepts and Methods (Kluwer, Dordrecht, 1995)

[2] A. Acín, N. Gisin, Ll. Masanes; *From Bell's Theorem to Secure Quantum Key Distribution* , Phys. Rev. Lett. 97, 120405 (2006)

[3] Antonio Acín, Lluis Masanes, Nicolas Gisin; *Equivalence between two-qubit entanglement and secure key distribution*, Phys. Rev. Lett. 91, 167901 (2003)

[4] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu and D. Roberts, *Nonlocal correlations as an information-theoretic resource*, Phys. Rev. A. **71**, 022101 (2005).

[5] J. Barrett, L. Hardy and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).

[6] harles H. Bennett, Gilles Brassard i Artur K. Ekert *Criptografía cuántica* Temas 10 Investigación y Ciencia

[7] C.H. Bennett and G. Brassard, Proceedings IEEE Int. Conf. on Computers, Systems and Signal Processing, Bangalore, India (IEEE, New York, 1984), pp. 175-179.

[8] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York 1991.

[9] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[10] J. S. Bell, Physics **1**, 195 (1964).

[11] M.A. Nielsen, I.L. Chuang *Quantum Computation and Quantum Information* Cambridge University Press

[12] U. M. Maurer, *Secret Key Agreement by Public Discussion from Common Information* IEEE Trans. Inf. Theory **39**, 733 (1993).

[13] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).

[14] U. M. Maurer, S. Wolf ;*Unconditionally Secure Key Agreement and the Intrinsic Conditional Information* IEEE Trans. Inf. Theory **45**, 499 (1999).

[15] Ll. Masanes, A. Acín, N. Gisin; *General properties of nonsignaling theories*,Physical Review A **73**, 012112 (2006)

[16] Ll. Masanes; *Necessary and sufficient condition for quantum-generated correlations*, quant-ph/0309137

[17] Navascues, M., S. Pironio, A. Acín, New J. Phys. **10**, 073013 (2008)

[18] V. Scarani, N. Gisin, N. Brunner, Ll. Masanes, A. Acín, S. Pino; *Secrecy extraction from non-local correlations*, Phys. Rev. A 74, 042339 (2006)

[19] C. Shannon *A Mathematical Theory of Communication*

[20] C. E. Shannon; *Communication Theory of Secrecy Systems* Bell Syst. Tech. J., vol 28, pp. 656-715, Oct 1949.

[21] D. Collins, N. Gisin, N. Linden, S. Massar, S. Popescu, *Bell inequalities for arbitrarily high dimensional systems* Phys. Rev. Lett. **88**, 040404 (2002)

[22] N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Rev. Mod. Phys **74**, 145 (2002)

[23] M. Dušek, N. Lütkenhaus, M. Hendrych, quant-ph/0601207; to appear in: Progress in Optics, vol. 49, Edt. E. Wolf (Elsevier)

[24] B.S. Tsirelson, Hadronic J. Supplement **8**, 329 (1993)

[25] R.F. Werner, M.M. Wolf, Phys. Rev. A **64**, 032112 (2001), paragraph V.C, and references therein.

[26] M. Curty, M. Lewenstein, N. Lütkenhaus, Phys. Rev. Lett. **92**, 217903 (2004)

[27] A. Acín, N. Gisin, Phys. Rev. Lett. **94**, 020501 (2005)

[28] D. Collins, N. Gisin, J. Phys. A: Math. Gen. **37** 1775 (2004)

[29] J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, Phys. Rev. Lett. **23**, 880 (1969)

[30] B.S. Cirel'son, Lett. Math. Phys. **4**, 83 (1980)

[31] Feller, Probability theory

[32] I. Csiszár, J. Körner, IEEE Trans. Inf. Theory **24**, 339 (1978)

[33] M. Christandl, R. Renner and S. Wolf, Proc. ISIT 2003, 258, Yokohama, Japan.

[34] R. Renner and S. Wolf, Adv. in Crypt., EUROCRYPT '03, Lecture Notes in Computer Science, Springer-Verlag, 2003.

[35] N. Gisin and S. Wolf, *Proceedings of CRYPTO 2000*, Lecture Notes in Computer Science **1880**, 482, Springer-Verlag, 2000, quant-ph/0005042.

[36] A. Acín, J. I. Cirac and L. Masanes, Phys. Rev. Lett. **92**, 107903 (2004); L. Masanes and A. Acín, cs.CR/0501008, accepted for publication in IEEE Trans. Inf. Theory.

[37] L. Masanes, Quant. Inf. Comput. **3**, 345 (2002)

[38] L. Masanes, A. Winter, quant-ph/0606049

[39] A. Acín, S. Massar, S. Pironio, quant-ph/0605246

[40] B. Kraus, N. Gisin and R. Renner, Phys. Rev. Lett. **95**, 080501 (2005).

[41] R. Renner, *Security of Quantum Key Distribution*, PhD thesis, quant-ph/0512258.

[42] I. Devetak and A. Winter, Proc. R. Soc. Lond. A **461**, 207 (2005).

[43] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar and V. Scarani *Device-independent quantum key distribution secure against collective attacks* New Journal of Physics. 11(4):045021, (2009).

[44] R. Arnon-Friedman, R. Renner and T. Vidick, quant-ph/1607.01797v1. (2016).