Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

PROJECTE FINAL DE CARRERA

QUANTUM TECHNOLOGIES FOR INFORMATION PROCESSING

Estudis: Enginyeria de Telecomunicació

Autor: Jordi Tura i Brugués

Director/a: Sebastià Xambó i Descamps
Co-director: Antonio Acín dal Maschio
Ponent ETSETB: Juan P. Torres

Any: 2010-2011

Universitat Politècnica de Catalunya

Escola Tècnica Superior d'Enginyeria de Telecomunicacions de Barcelona

Projecte Final de Carrera

Quantum Technologies for Information Processing

Jordi Tura i Brugués

Advisor: Sebastià Xambó Descamps
Co-Advisor: Antonio Acín dal Maschio
ETSETB Proponent: Juan Pérez Torres

Teoria del Senyal i Comunicacions
Matemàtica Aplicada II

Abstract

Keywords: Quantum, information, processing, computation, coding, cryptography, entanglement, field theory, entropy, key distribution, protocols.

Since its genesis, quantum mechanics has proved to be a very accurate model for predicting the behavior of the world below the nanoscale. However, crucial breakthroughs in technology were needed in order to be able to effectively access and manipulate such small magnitudes.

During the last twenty years, the field of quantum information processing has experienced a growing interest, in its many variants, both theoretically and practically. Despite being still at a very basic stage, expectations are high.

The uniqueness of quantum phenomena (superposition of states, creation of entanglement, etc.) have no classical analogue and allow novelties such as another paradigm of computation, more secure communications, quantum teleportation, quantum dense coding, etc. which are presented and analyzed here.

The aim of this Thesis is to present in a unified way the main mathematical methods used in quantum information processing, as well as the state of the art of their corresponding technological implementations.

Our contributions are based in making a self-contained presentation; seeking completeness, that is, treating the most relevant fields of research involved, focusing on their relations; and picking the most relevant, insightful references, given the quantity of literature produced in this field. We also discuss some of the fundamental questions that remain still unanswered, and which are the current lines of research to shed light on them.

Contents

Introduction

Since its genesis, quantum mechanics has been a very accurate model for the description of reality below the nanoscale: molecules, atoms and subatomic particles behave according to quantum physics in a very exact manner. Probability also plays a main role in quantum mechanics and some extraordinary, counter-intuitive phenomena arise when one considers a world described by quantum physics. For example, an atom can be in a superposition of several energy levels, a single photon can be traveling on two different paths at the same time; a particle can pass through a potential barrier with non-zero probability (tunnel effect); a measure performed on a quantum system, in general, disturbs it, etc. But, perhaps, the quintessential quantum effect would be entanglement: a measure performed on a system can affect the state of another system, regardless of their distance. These effects have no classical analogue. In fact, classical physics, which comprises any theory that is not quantum, including relativity, is deterministic, and it has been proved to be highly accurate at larger scales.

Yet this convenient partitioning of the world between quantum and classical could be a myth. Some physicists defend that the world is quantum at all scales, and classical physics is just a useful approximation: the world looks classical because the complex interactions an object endures with its environment tend to conceal quantum effects from our view. Over the past several years, experimentalists have observed quantum effects in a growing number of macroscopic systems[1].

In the last 20 years, technological breakthroughs have made it possible to make use of these unique phenomena. Despite still being at an embryonic stage, technology will soon be endowed with new tools and resources, due to the possibility of constructing, in the near future, physical realizations of systems which exploit quantum features. Being developed within a rich interdisciplinary context, from as varied fields as theoretical and experimental quantum mechanics, quantum field theory, information theory, computer science or cryptography, it has become a new and active field of interest.

The vast quantity of literature available nowadays has required an effort to summarize, as well as the need to pick up the most relevant bibliographical sources: Classical references, such as [**48, 53**], or more recent publications, such as [**5, 7, 8, 36**], representing the state of the art of the current technology, are the sources that have been most widely consulted. Each chapter

[1] For example, the prestigious magazine Scientific American features this month at front cover the article *Living in a Quantum World*: Experimental evidence, from the entanglement created among the 10^{20} atoms of a grain of salt, to the achievement of a mechanical springboard of 40 microns long that oscillates at two frequencies at once (a *huge* quantum harmonic oscillator) is given. Moreover, quantum effects need not be created in a laboratory: European robins (*Erithacus rubecula*) have a molecule that contains two entangled electrons with total zero spin; when it absorbs light the electrons separate and become susceptible to external influences such as Earth magnetic field. This could be the secret of the extraordinary sense of orientation such birds possess.

contains a brief introduction which includes general reference sources, whereas sources appearing in a specific section concern only that part.

Although one encounters high similarity among works concerning the same subject, we have oriented the exposition in a comprehensive, yet straightforward manner. Some parts are treated in a *top-down* scheme: we present the model and then we analyze it, as one usually would do in mathematics; but some other parts have required a *bottom-up* analysis in order to derive the model and justify its choice, which is a more physical or technological approach.

The aim of this project is to present in a unified way the main mathematical methods used in quantum information processing, as well as the state of the art of its corresponding technological implementations. The contribution we pursue is the interconnection within our reach of all the different subjects we treat here, to make an effort to relate the different fields quantum information processing comprises and to give a general overview of where currently are the frontiers of quantum information processing, both theoretical and practical.

This project has been organized in the following parts, in order to treat the subject in the most self-contained possible way:

In the first part, we introduce the mathematical concepts which are needed to describe quantum systems, as a foundation for the remaining chapters. These subjects include Hilbert spaces and classical information theory.

In the second part, we focus on the physical concepts that are used to describe quantum mechanics. We begin with introducing the postulates, first with an informal discussion and then with an axiomatic description. We move to quantum probability, and analyze why classical probability does not suffice, as well as the consequences it implies: some relevant impossibilities and novelties due to quantum effects. Finally, we end this part with an introduction to quantum field theory, as a bridge from the mathematical model to the experiments that we discuss in the last part.

The third part of the project contains four important areas of quantum information processing. We begin with quantum computation, continue with quantum entropy and quantum information, move to quantum cryptography and end with quantum coding. At the end of each of these chapters we give a broad view of the current research lines: In which algorithms can a quantum computer outperform a classical one; the problem of characterization of entanglement; the possibility of certifying a sequence of numbers is random via quantum correlations; and the use of degenerate quantum error correcting codes, which need not completely reveal the error syndrome in order to perform a correction.

The last part focuses on some of the physical technologies that are used today to manipulate quantum bits and produce interaction between them. We highlight quantum optics, cavity quantum electrodynamics and ion traps, together with the Jaynes-Cummings model to analyze them. At the end, we present the realization of the Cirac-Zoller CNOT gate, which creates entanglement between two qubits and represents a milestone in the field of quantum information processing.

Finally, we present our conclusions and treat the future lines of development that may exist.

Part 1

Mathematical Preliminaries

Chapter 1
Hilbert Spaces

In this chapter we will present the backbone of quantum physics, which is provided by the theory of linear operators acting on Hilbert Spaces [**8, 48, 51**].

To begin with the formalism of quantum mechanics, one has to consider the Hilbert space.

Definition 1.1. A Hilbert Space is an inner product space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ which is complete with respect to the norm induced by $\langle\cdot,\cdot\rangle$.

In the context of quantum computing, we consider $\mathcal{H}$ as a $\mathbb{C}$ -vector space. The completeness is required to ensure the convergence of some linear operators acting on $\mathcal{H}$, such as the Quantum Fourier Transform.

Remark 1.1. Nevertheless, we will mostly consider finite dimension $\dim_{\mathbb{C}}\mathcal{H} < \infty$, as this is sufficient for Quantum Information Processing purposes.

In the following chapters, we will also follow Dirac's notation: The vectors of $\mathcal{H}$ will be denoted by $|v\rangle$.

1.1. Orthogonal expansions

Let $\mathcal{H}$ be a complex vector space. For any $\lambda \in \mathbb{C}$, we will denote its conjugate as λ^* and its modulus as $|\lambda|$.

Definition 1.2. A map $\langle\cdot,\cdot\rangle : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{C}$ is called (Hermitian) inner product if he following properties hold:

(1) $\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle + \langle \boldsymbol{y}, \boldsymbol{z}\rangle \ \ \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{H}$,
(2) $\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = \lambda^*\langle \boldsymbol{x}, \boldsymbol{y}\rangle \ \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}, \lambda \in \mathbb{C}$,
(3) $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = (\langle \boldsymbol{y}, \boldsymbol{x}\rangle)^* \ \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$
(4) $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0 \ \ \forall \boldsymbol{x} \in \mathcal{H}$ and $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0 \Rightarrow \boldsymbol{x} = 0$.

In other words, properties 1-2 say that $\langle\cdot,\cdot\rangle$ is a sesquilienar form, property 3 says that it is Hermitian and property 4 that it is positive-definite.

These conditions imply the *Cauchy-Schwartz* inequality $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle|^2 \leq \langle \boldsymbol{x}, \boldsymbol{x}\rangle \cdot \langle \boldsymbol{y}, \boldsymbol{y}\rangle$, which is the particular case $p = q = 2$ of the Hölder inequality $\|\boldsymbol{x}\boldsymbol{y}\|_1 \leq \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q$.

The inner product $\langle\cdot,\cdot\rangle$ also induces a norm $\|\boldsymbol{x}\| := \sqrt{\langle\boldsymbol{x},\boldsymbol{x}\rangle}\ \ \forall\boldsymbol{x}\in\mathcal{H}$, which satisfies the triangle inequality $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\|+\|\boldsymbol{y}\|\ \ \forall\boldsymbol{x},\boldsymbol{y}\in\mathcal{H}$, which is the special case $p=2$ of the Minkowski inequality $\|\boldsymbol{x}+\boldsymbol{y}\|_p \leq \|\boldsymbol{x}\|_p+\|\boldsymbol{y}\|_p$.

Finally, the condition of completeness is that every Cauchy sequence in $\mathcal{H}$ must be convergent.

Definition 1.3. Two vectors $\boldsymbol{x},\boldsymbol{y}\in\mathcal{H}$ are called orthogonal if $\langle\boldsymbol{x},\boldsymbol{y}\rangle=0$ and it will be denoted $\boldsymbol{x}\perp\boldsymbol{y}$.

For any subset $H\subset\mathcal{H}$, the orthogonal subset is defined as

$$H^{\perp} := \{\boldsymbol{x}\in\mathcal{H} : \boldsymbol{x}\perp\boldsymbol{h}\ \forall\boldsymbol{h}\in H\},$$

and it is a closed subspace.

Theorem 1.1. Let $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots$ be an orthonormal basis in a Hilbert space $\mathcal{H}$. For any $\boldsymbol{x}\in\mathcal{H}$, the expansion

$$\boldsymbol{x}=\sum_n\langle\boldsymbol{x}_n,\boldsymbol{x}\rangle\boldsymbol{x}_n$$

holds.

Theorem 1.2. *Projection theorem.* Let $\mathcal{M}$ be a closed subspace of $\mathcal{H}$. Any vector $\boldsymbol{x}\in\mathcal{H}$ can be written in a unique way in the form $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{y}$, where $\boldsymbol{x}_0\in\mathcal{M}$ and $\boldsymbol{y}\in\mathcal{M}^{\perp}$.

Definition 1.4. In the context of the previous theorem, the map $P:\boldsymbol{x}\mapsto\boldsymbol{x}_0$ is called the orthogonal projection onto the subspace $\mathcal{M}$.

P is linear and satisfies $P^2=P$.

Definition 1.5. Let $\mathcal{H}$, $\mathcal{K}$ be Hilbert spaces. The norm of a linear operator $A:\mathcal{H}\longrightarrow\mathcal{K}$ is defined as

$$\|A\| := \sup\{\|Ax\| : \boldsymbol{x}\in\mathcal{H}, \|\boldsymbol{x}\|=1\}.$$

If $\|A\|<\infty$, we say that A is bounded.

1.2. The adjoint operator

Definition 1.6. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces. If $T:\mathcal{H}\longrightarrow\mathcal{K}$ is a bounded linear operator, its adjoint is the only operator $T^{\dagger}:\mathcal{K}\longrightarrow\mathcal{H}$ which satisfies

$$\langle\boldsymbol{x},T\boldsymbol{y}\rangle_{\mathcal{K}}=\langle T^{\dagger}\boldsymbol{x},\boldsymbol{y}\rangle_{\mathcal{H}}\ \ \forall\boldsymbol{x}\in\mathcal{K},\boldsymbol{y}\in\mathcal{H}.$$

Definition 1.7. We will denote by $\mathcal{B}(\mathcal{H})$ the set of all bounded linear operators $T:\mathcal{H}\longrightarrow\mathcal{H}$.

$T\in\mathcal{B}(\mathcal{H})$ is called self-adjoint if $T^{\dagger}=T$.

This definition is equivalent to $\langle\boldsymbol{x},T\boldsymbol{x}\rangle\in\mathbb{R}\ \ \forall\boldsymbol{x}\in\mathcal{H}$.

Example 1.1. Any orthogonal projection is self-adjoint.

Properties 1.1. Let A,B be bounded linear operators and $\lambda\in\mathbb{C}$. The adjoint has the following properties:

(1) $(A+B)^{\dagger}=A^{\dagger}+B^{\dagger},(\lambda A)^{\dagger}=\lambda^{*}A^{\dagger}$,
(2) $(A^{\dagger})^{\dagger}=A,(AB)^{\dagger}=B^{\dagger}A^{\dagger}$,
(3) if A is invertible, $(A^{-1})^{\dagger}=(A^{\dagger})^{-1}$,

(4) $\|A\| = \|A^\dagger\|$.

Definition 1.8. An invertible operator $U \in \mathcal{B}(\mathcal{H})$ is called unitary if $U^{-1} = U^\dagger$.

1.3. Tensor product of Hilbert Spaces and Operators

Definition 1.9. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces. Their algebraic tensor product consists of the formal finite sums

$$\sum_{i,j} \boldsymbol{x}_i \otimes \boldsymbol{y}_j \qquad \boldsymbol{x}_i \in \mathcal{H}, \boldsymbol{y}_j \in \mathcal{K}.$$

When $\mathcal{H}$ and $\mathcal{K}$ are finite-dimensional spaces, the above construction defines the tensor product Hilbert space $\mathcal{H} \otimes \mathcal{K}$; otherwise the algebraic tensor product should be completed in order to obtain a Hilbert space.

The following rules are used for computation:

- $(\boldsymbol{x}_1 + \boldsymbol{x}_2) \otimes \boldsymbol{y} = \boldsymbol{x}_1 \otimes \boldsymbol{y} + \boldsymbol{x}_2 \otimes \boldsymbol{y}$
- $\boldsymbol{x} \otimes (\boldsymbol{y}_1 + \boldsymbol{y}_2) = \boldsymbol{x} \otimes \boldsymbol{y}_1 + \boldsymbol{x} \otimes \boldsymbol{y}_2$
- $(\lambda \boldsymbol{x}) \otimes \boldsymbol{y} = \lambda(\boldsymbol{x} \otimes \boldsymbol{y})$
- $\boldsymbol{x} \otimes (\lambda \boldsymbol{y}) = \lambda(\boldsymbol{x} \otimes \boldsymbol{y})$

And the inner product of $\mathcal{H} \otimes \mathcal{K}$ acts as

$$\langle \sum_{i,j} \boldsymbol{x}_i \otimes \boldsymbol{y}_j, \sum_{k,l} \boldsymbol{z}_k \otimes \boldsymbol{w}_l \rangle_{\mathcal{H} \otimes \mathcal{K}} = \sum_{i,j,k,l} \langle \boldsymbol{x}_i, \boldsymbol{z}_k \rangle_{\mathcal{H}} \cdot \langle \boldsymbol{y}_j, \boldsymbol{w}_l \rangle_{\mathcal{K}}.$$

The definition of the tensor product of several Hilbert spaces extends similarly.

Proposition 1.1. If $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots\}$ and $\{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots\}$ are bases in $\mathcal{H}$ and $\mathcal{K}$, respectively, then $\{\boldsymbol{e}_i \otimes \boldsymbol{f}_j\}_{i,j}$ is a basis in the tensor product space. This shows that

$$\dim(\mathcal{H} \otimes \mathcal{K}) = \dim(\mathcal{H}) \cdot \dim(\mathcal{K}).$$

Notation 1.1. Let $\{\boldsymbol{e}_1, \ldots \boldsymbol{e}_k\}$ be a basis in $\mathcal{H}$ and $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_l\}$ be a basis in $\mathcal{K}$.

If (A_{ij}) is the matrix $A \in \mathcal{B}(\mathcal{H})$ and $(B_{i'j'})$ is the matrix $B \in \mathcal{B}(\mathcal{K})$, then $A \otimes B \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ acts as

$$(A \otimes B)(\boldsymbol{e}_j \otimes \boldsymbol{f}_{j'}) = \sum_{i,i'} A_{ij} B_{i'j'} \boldsymbol{e}_i \otimes \boldsymbol{f}_{i'}.$$

We will order the tensor product basis lexicographically:

$\{\boldsymbol{e}_1 \otimes \boldsymbol{f}_1, \ldots, \boldsymbol{e}_1 \otimes \boldsymbol{f}_l, \boldsymbol{e}_2 \otimes \boldsymbol{f}_1, \ldots, \boldsymbol{e}_2 \otimes \boldsymbol{f}_l, \ldots, \boldsymbol{e}_k \otimes \boldsymbol{f}_1, \ldots, \boldsymbol{e}_k \otimes \boldsymbol{f}_l\}$. With this ordering fixed, the matrix of $A \otimes B$ can be written as (the Kronecker product of $A \otimes B$)

$$A \otimes B = \begin{pmatrix} A_{11}B_{11} & \cdots & A_{11}B_{1l} & & A_{1k}B_{11} & \cdots & A_{1k}B_{1l} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ A_{11}B_{l1} & \cdots & A_{11}B_{ll} & & A_{1k}B_{l1} & \cdots & A_{1k}B_{ll} \\ & \vdots & & \ddots & & \vdots & \\ A_{k1}B_{11} & \cdots & A_{k1}B_{1l} & & A_{kk}B_{11} & \cdots & A_{kk}B_{1l} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ A_{k1}B_{l1} & \cdots & A_{k1}B_{ll} & & A_{kk}B_{l1} & \cdots & A_{kk}B_{ll} \end{pmatrix},$$

or, in shorter notation,

$$A \otimes B = \begin{pmatrix} A_{11} \cdot B & \cdots & A_{1k} \cdot B \\ \vdots & \ddots & \vdots \\ A_{k1} \cdot B & \cdots & A_{kk} \cdot B \end{pmatrix}$$

Notation 1.2. Let $\mathcal{H}$ be a Hilbert space. The k-fold tensor product $\mathcal{H} \otimes \cdots \otimes \mathcal{H}$ is called the kth tensor power of $\mathcal{H}$, and noted $\mathcal{H}^{\otimes k}$. If $\{A^{(i)}\}_i \subset \mathcal{B}(\mathcal{H})$, then $A^{(1)} \otimes \cdots \otimes A^{(k)}$ is a linear transformation on $\mathcal{H}^{\otimes k}$; if all A's are the same, it is denoted $A^{\otimes k}$.

1.4. Positive Operators

Definition 1.10. The spectrum $Spec(T)$ of an operator $T \in \mathcal{B}(\mathcal{H})$ consists of all the numbers $\lambda \in \mathbb{C}$ such that the operator $\lambda\mathbb{I} - T$ does not have a bounded inverse.

If $\dim_{\mathbb{C}} \mathcal{H} < \infty$, the definition is equivalent to

$$Spec(T) = \{\lambda \in \mathbb{C} \,: \exists \boldsymbol{x} \in \mathcal{H}, \boldsymbol{x} \neq 0, \lambda\boldsymbol{x} - T\boldsymbol{x} = 0\}.$$

In this case, we say that $\boldsymbol{x}$ is an eigenvector and λ its corresponding eigenvalue.

Remark 1.2. If A is a self-adjoint matrix, its eigenvalues are real and the eigenvectors corresponding to different eigenvalues are orthogonal.

As such, A can be written into the form

$$A = \sum_{i=1}^{k} \lambda_i E_i,$$

where $\lambda_1, \ldots, \lambda_k$ are the different eigenvalues and E_i is the orthogonal projection onto the subspace spanned by the corresponding eigenvalue λ_i, $1 \leq i \leq k$.

Definition 1.11. $T \in \mathcal{B}(\mathcal{H})$ is positive if $\langle \boldsymbol{x}, T\boldsymbol{x} \rangle \geq 0 \ \ \forall \boldsymbol{x} \in \mathcal{H}$. We will simply write $T \geq 0$.

Remark 1.3. A positive operator is self-adjoint.

Theorem 1.3. (Sylvester criterion) Let $T \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ be a basis in the Hilbert space $\mathcal{H}$. T is positive if, and only if, for any $1 \leq k \leq n$, the determinant of the $k \times k$ matrix

$$(\langle \boldsymbol{e}_i, T\boldsymbol{e}_j \rangle)_{i,j=1}^{k}$$

is positive.

Remark 1.4. The spectrum of a non-negative operator lies in $\mathbb{R}^+$. In particular, its eigenvalues are non-negative. Conversely, if all the eigenvalues of a non-negative (self-adjoint) operator acting on a finite dimensional Hilbert space lie in $\mathbb{R}^+$, then it is non-negative.

Notation 1.3. Positive matrices are also called positive semi-definite.

Definition 1.12. Let $T \in \mathcal{B}(\mathcal{H})$ and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ be a basis in the Hilbert space $\mathcal{H}$. The trace of $\mathcal{H}$ is the sum of the elements of its diagonal:

$$Tr(T) = \sum_{i=1}^{n} \langle \boldsymbol{e}_i, T\boldsymbol{e}_i \rangle.$$

Remark 1.5. The trace is the sum of the eigenvalues and is independent of the basis chosen to compute it. It also satisfies the cyclicity property

$$Tr(AB) = Tr(BA) \quad \forall A, B \in \mathcal{B}(\mathcal{H}). \tag{1.1}$$

1.5. Spectral Theorem

In this section we extend the results of the previous one to an arbitrary self-adjoint operator A. In this general situation, the spectra needs not to be discrete.

Definition 1.13. More formally, let $\mathcal{X}$ be a complete, separable, metric space and let $\mathcal{H}$ be a Hilbert space. We consider the smallest σ-algebra which contains all the $\mathcal{X}$-open subsets; its measurable [**28**] subsets are called Borel subsets.

Now, let's assume that for each Borel subset $B \subset \mathcal{X}$, we can find a positive operator $E(B) \in \mathcal{B}(\mathcal{H})$ which satisfies

(1) $0 \leq E(B) \leq \mathbb{I}, \quad E(\emptyset) = 0, \quad E(\mathcal{X}) = \mathbb{I}.$
(2) If $\{B_i\}_i$ is a sequence of pairwise disjoint Borel subsets of $\mathcal{X}$, and $B = \cup_{i=1}^{\infty} B_i$, then for every $\boldsymbol{e} \in \mathcal{H}$,

$$(E(B))(\boldsymbol{e}) = \left(\sum_{i=1}^{\infty} E(B_i) \right) (\boldsymbol{e}).$$

In this case, E is called a positive operator-valued measure, which we shall denote POVM.

Remark 1.6. The comparison $A \leq B$ between operators should be taken as the difference operator being positive: $B - A \geq 0$.

Remark 1.7. We assume that the series $\left(\sum_{i=1}^{\infty} E(B_i) \right)$ is an operator which exists and is in $\mathcal{B}(\mathcal{H})$.

Remark 1.8. The most important cases are when $\mathcal{X}$ is a finite set, $\mathbb{R}$, or the unit circle $S^1 \subset \mathbb{C}$.

Let's consider a function $f : \mathcal{X} \longrightarrow \mathbb{C}$, which we want to integrate, with respect to a POVM E on $\mathcal{X}$.

For example, in the case $\#\mathcal{X} < \infty$,

$$\int_{\mathcal{X}} f(x) dE(x) = \sum_{x \in \mathcal{X}} f(x) E(\{x\}),$$

which is a finite sum.

In the general case, the definition of the integral can be reduced to many integrals with respect to more common measures. More precisely, given a vector $\boldsymbol{e} \in \mathcal{H}$,

$$\mu_{\boldsymbol{e}}(B) = \langle \boldsymbol{e}, (E(B))(\boldsymbol{e}) \rangle$$

gives a positive measure on the Borel sets of $\mathcal{X}$.

Definition 1.14. We say that the integral

$$\int_{\mathcal{X}} f(x) dE(x) = T \in \mathcal{B}(\mathcal{H})$$

if $\langle \boldsymbol{e}, T\boldsymbol{e} \rangle = \int_{\mathcal{X}} f(x) d\mu_e(x)$ holds $\forall \boldsymbol{e} \in \mathcal{H}$.

Definition 1.15. A POVM E is called a projection-valued measure if $E(B)$ is a projection operator for every Borel set B; that is, $E(B) = E(B) \circ E(B) = E(B)^2$.

Theorem 1.4. *Spectral Theorem* for a bounded self-adjoint operator.
Let $A = A^\dagger \in \mathcal{B}(\mathcal{H})$. Then, there exists a unique projection-valued measure E on the real line $\mathbb{R}$, such that

$$A = \int \lambda dE(\lambda).$$

Moreover, if $B \subset \mathbb{R}$ and the spectrum of A is disjoint, then $E(B) = 0$ and for every continuous function f defined on the spectrum of A we have

$$f(A) = \int f(\lambda) dE(\lambda).$$

Remark 1.9. The projection-valued measure in the theorem is called the spectral measure of the operator A. A similar result holds for unbounded self-adjoint operator A, but in this case A and $f(A)$ are not defined everywhere. Also, a similar result holds in the case of unitary operators; then the spectral measure is on the unit circle.

1.6. Schmidt Decomposition

In this section, we shall demonstrate the existence of a very useful decomposition, known as the Schmidt decomposition.

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces.

Theorem 1.5. Schmidt decomposition theorem.
For any vector $\boldsymbol{\psi} \in \mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, there exist orthonormal states $\{\boldsymbol{e}_i^1\}_i \subset \mathcal{H}_1$ and $\{\boldsymbol{e}_i^2\}_i \subset \mathcal{H}_2$ such that

$$\boldsymbol{\psi} = \sum_{i=1}^{k} \sqrt{p_i} \boldsymbol{e}_i^1 \otimes \boldsymbol{e}_i^2,$$

with $p_i \in \mathbb{R}^+$ satisfying the condition $\sum_{i=1}^{k} p_i = 1$.

Remark 1.10. The states $\boldsymbol{e}_i^1$ and $\boldsymbol{e}_i^2$ depend on the particular state $\boldsymbol{\psi}$ that we wish to expand.

Remark 1.11. The Schmidt decomposition cannot be extended to tensor product spaces which consist of more than 2 components.

Chapter 2
Classical Information Theory

In this chapter we will present the fundamentals of classical information theory: the description of entropy, Shannon's theorem for noiseless source coding and noisy channel coding. We will also briefly discuss how far we are from reaching the theoretical limits imposed by these theorems [**42, 48, 58, 66, 67, 71**].

For the introduction of Shannon's Theorems, we will follow the exposition of [**66**]. We also give some proofs, with the objective of showing how one manipulates the quantities that are defined in order to announce such theorems.

2.1. Entropy

Definition 2.1. The information gained by knowing that an event A has occurred is defined by

$$\iota(A) = -\log_2 \mathbf{P}(A),$$

where $\mathbf{P}$ stands for the underlying probability distribution. The information is measured in bits.

Remark 2.1. A dual point of view of the definition is that $\iota(A)$ is the amount of uncertainty before learning that event A has occurred. We can also say that information measures the necessary amount of data needed to specify event A.

Definition 2.2. Let X be a discrete random variable which takes finitely many distinct values $x_1, \ldots, x_m$ with probabilities $p_1, \ldots, p_m$, where $p_i = \mathbf{P}(X = x_i)$. The (Shannon) entropy of X is the expected amount of information gained when learning the value of X:

$$h(X) = -\sum_{i=1}^{m} p_i \log_2 p_i.$$

Remark 2.2. The entropy depends on the probability distribution $\mathbf{P}$, but not on the values of X.

Remark 2.3. If some of the $p_i = 0$ we take the convention $p_i \log_2 p_i = 0$ (by continuity, $\lim_{x \to 0^+} -x \log_2 x = 0$).

Definition 2.3. Let X be a continuous random variable with a probability density function $p(x)$. The entropy is defined in this case by an integral, not a sum. More precisely,

$$h(X) = -\int p(x) \log_2 p(x) dx$$

Remark 2.4. Unlike the entropy of a discrete random variable, the entropy of a continuous one may take negative values, since the density probability function needs not take values in the interval $[0,1]$.

The relative entropy is a very useful measure to determine how close are two probability distributions, which we shall denote $p_X(x), p_Y(x)$, taking values on the same set.

Definition 2.4. If $p_Y(x) = \mathbf{P}(Y = x) > 0 \ \forall x$, then the relative entropy $h(X||Y)$ can be defined:

$$h(X||Y) = -\sum_x p_X(x) \log_2 \frac{p_X(x)}{p_Y(x)}.$$

The following proposition should motivate the fact that the relative entropy can be seen as some kind of (asymmetric) distance measure [**48**]:

Theorem 2.1. *Non-negativity of the relative entropy*:
The relative entropy is non-negative; i.e.,

$$h(X||Y) \geq 0.$$

The equality holds if, and only if,

$$p_X(x) = p_Y(x) \quad \forall x.$$

Remark 2.5. The relative entropy is useful since other entropic quantities can be derived from it, as special cases.

Example 2.1. For example, if X takes values on a set of d elements, it follows that

$$0 \leq h(X) \leq \log_2 d.$$

Indeed, consider $p_X(x)$ a probability distribution for X and let Y be a uniform random variable, so that $p_Y(x) = \frac{1}{d}$ for all x. Then

$$h(X||Y) = \sum_x p_X(x) \log_2 \frac{p_X(x)}{1/d} = -h(X) - \sum_x p_X(x) \log_2(1/d) = \log_2 d - h(X) \geq 0.$$

Note that $h(X) = \log_2 d$ if, and only if, X is uniformly distributed.

Let X and Y be random variables taking values x_i, y_j respectively. We denote its joint probability distribution $p_{X,Y}(x_i, y_j) = \mathbf{P}(X = x_i, Y = y_j)$ and its conditional probability distribution $p_{X|Y}(x_i, y_j) = \mathbf{P}(X = x_i | Y = y_j)$. We want to relate the information content of X to the information content of Y. The following two quantities help us find their relation:

Definition 2.5. Let X, Y be random variables taking values x_i, y_j.
The joint entropy $h(X, Y)$ is defined as

$$h(X, Y) = -\sum_{x_i, y_j} p_{X,Y}(x_i, y_j) \log_2 p_{X,Y}(x_i, y_j).$$

The conditional entropy $h(X|Y)$ of X, given Y, is defined as

$$h(X|Y) = -\sum_{x_i, y_j} p_{X,Y}(x_i, y_j) \log_2 p_{X|Y}(x_i, y_j).$$

Remark 2.6. To illustrate Definition 2.5, let us suppose that we know the value of Y, having gained $h(Y)$ bits of information about the pair (X, Y). The remaining uncertainty about the pair (X, Y) is associated with the remain lack of knowledge about X, even though we know Y. Note that the equivalent definition would be

$$h(X|Y) = h(X, Y) - h(Y),$$

which is, indeed, satisfied.

Another quantity, which measures how much information X and Y have in common is given in Definition 2.6. To understand it, suppose we add the information content of X, $h(X)$ to Y. Information which is common between them will be counted twice, and information not common only once. The mutual information X and Y have in common is then $h(X) + h(Y) - h(X, Y)$.

Definition 2.6. The mutual information between X and Y is defined as

$$\iota(X : Y) = \sum_{x,y} p_{X,Y}(x, y) \log_2 \frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)} = h(X) + h(Y) - h(X, Y)$$

Figure 2.1 shows the relationship between these quantities.

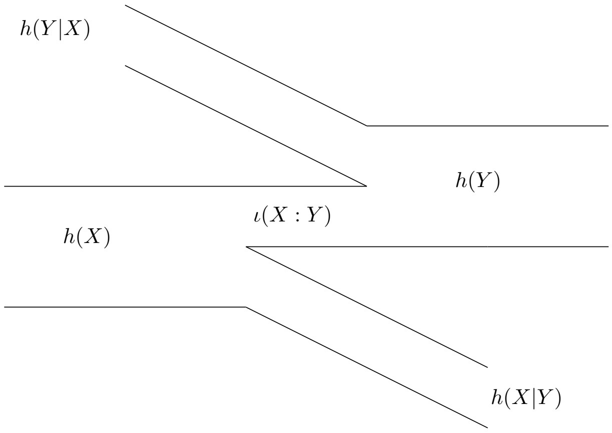

FIG. 2.1. Representation of the relationship between different entropic quantities.

2.2. Source Coding

In order to be able to compress data in an efficient manner, the main idea is to identify relatively small sets which occur with higher probability [**66**].

Definition 2.7. For a given $R > 0$, a source generating random strings $\mathcal{U}^{(n)} = (U_1, \ldots, U_n)$, where the symbols $U_i \in I$ are taken from an alphabet I, we say that this source is reliably encodable at rate R if, for any n, there exists a set $A_n \subseteq I^n$ such that

$$\#A_n \leq 2^{nR} \quad \text{and} \quad \lim_{n\to\infty} \mathbf{P}(\mathcal{U}^{(n)} \in A_n) = 1.$$

Remark 2.7. The idea of this definition is that we can label the members of the set A_n with a string of length not greater than nR. Since they are *nearly always* in A_n, the average length of a compressed string will be close below nR.

We use this definition to introduce the concept of information rate:

Definition 2.8. The information rate H of a given source is the smallest reliable encoding rate:

$$H = \inf\{R : R \text{ is reliable}\}.$$

The information rate always depends on the size of the alphabet, as the following proposition [**66**] shows:

Proposition 2.1. For a source of alphabet size m,

$$0 \leq H \leq \log_2 m,$$

both bounds being attainable.

Remark 2.8. The left-hand equality is attained for a source the outputs of which are always the same. The right-hand equality is attained only for a source with uniformly, independent, identically distributed (IID) outputs: in any other case, for any set of strings A_n, $\mathbf{P}(A_n) = (1/m^n)\#A_n$, which goes to zero when $n \to \infty$ if $\#A_n \leq 2^{nR}$ and $R < \log_2 m$.

Remark 2.9. The quantity H is useful; nevertheless it is hard to compute. Introducing quantities D_n and ξ_n, we will see how H depends on D_n and D_n depends on ξ_n, which is easier to find, enabling us to calculate H.

Definition 2.9. The subset maximum is

$$D_n(R) = \max_{A:\ \#A \leq 2^{nR}} \mathbf{P}(\mathcal{U}^{(n)} \in A).$$

A set A on which the maximum is attained will be called a typical subset.
The log-likelihood per source letter is

$$\xi_n(\boldsymbol{u}^{(n)}) = -\frac{1}{n} \log_+ p_n(\boldsymbol{u}^{(n)}),$$

where $p_n(\boldsymbol{u}^{(n)}) = \mathbf{P}(\mathcal{U}^{(n)} = \boldsymbol{u}^{(n)})$ and $\log_+(x) = \begin{cases} \log_2(x) & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$.
With $\boldsymbol{u}^{(n)}$ we denote an observation of $\mathcal{U}^n$.

We need the following two lemmas [**66**] in order to announce Shannon's first coding theorem:

Lemma 2.1. $\forall \varepsilon > 0$, the information rate H satisfies

$$\lim_{n\to\infty} D_n(H + \varepsilon) = 1.$$

Moreover, if $H > 0$, then $D_n(H - \varepsilon) \not\to_n 1$.

Proof: $R = H + \varepsilon$ is a reliable encoding rate, by Definition 2.8. So, there exists a sequence of sets $A_n \subset I^n$, with $\#A_n \leq 2^{nR}$ and $\lim\limits_{n\to\infty} \mathbf{P}(\mathcal{U}^{(n)} \in A_n) = 1$. Definition 2.9 assures $D_n(R) \geq \mathbf{P}(\mathcal{U}^{(n)} \in A_n) \Rightarrow \lim\limits_{n\to\infty} D_n(R) = 1$.
On the other hand, if $H > 0$, we can find an ε such that $H - \varepsilon > 0$, but with $H - \varepsilon$ not reliable, by Definition 2.8. We choose C_n the typical subsets in Definition 2.9, which have cardinality

$\#C_n \leq 2^{nR}$; then, according to Definition 2.7, $\mathbf{P}(\mathcal{U}^{(n)} \in C_n) \not\to_n 1$ is the only remaining possibility. □.

Lemma 2.2. $\forall R, \forall \varepsilon > 0$,

$$\mathbf{P}(\xi_n \leq R) \leq D_n(R) \leq \mathbf{P}(\xi_n \leq R + \varepsilon) + 2^{-n\varepsilon}.$$

For a proof of this lemma, see [**66**].

Definition 2.10. A sequence of random variables $\{\eta_n\}_n$ converges in probability to a constant r if, $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}(|\eta_n - r| \geq \varepsilon) = 0.$$

Notation 2.1. Convergence in probability is denoted by $\eta_n \xrightarrow{\mathbf{P}} r$.

Theorem 2.2. (Shannon's first coding theorem)
If ξ_n converges in probability to a constant γ, then $\gamma = H$.

Proof: Let $\xi_n \xrightarrow{\mathbf{P}} \gamma$. Since $\xi_n \geq 0 \; \forall n$, its limit $\gamma \geq 0$. Then Lemma 2.2 assures that $\forall \varepsilon > 0$ we have $D_n(\gamma+\varepsilon) \geq \mathbf{P}(\xi_n \leq \gamma+\varepsilon) \geq \mathbf{P}(\gamma-\varepsilon \leq \xi_n \leq \gamma+\varepsilon) = \mathbf{P}(|\xi_n - \gamma| \leq \varepsilon) = 1 - \mathbf{P}(|\xi_n - \gamma| > \varepsilon)$. This means that $\lim_{n\to\infty} D_n(\gamma + \varepsilon) \geq 1$. Hence, $H \leq \gamma$, by Lemma 2.1.
In particular, $\gamma = 0 \Rightarrow H = 0$. If $\gamma > 0$, by the rightmost inequality in Lemma 2.2 (which holds for all R and $\varepsilon > 0$; in particular, for $\gamma - \varepsilon$ and $\varepsilon/2$, respectively): $D_n(\gamma - \varepsilon) \leq \mathbf{P}(\xi_n \leq \gamma - \varepsilon/2) + 2^{-n\varepsilon/2} \leq \mathbf{P}(|\xi_n - \gamma| \geq \varepsilon/2) + 2^{-n\varepsilon/2}$, which does not tend to 1. By Lemma 2.1, $H \geq \gamma$. □

Remark 2.10. Convergence $\xi_n \xrightarrow{\mathbf{P}} \gamma = H$ can be interpreted in the following way:

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbf{P}\left(2^{-n(H+\varepsilon)} \leq p_n(\boldsymbol{u}^{(n)}) \leq 2^{-n(H-\varepsilon)}\right) = 1,$$

which is called asymptotic equipartition property (AEP). In other words, for any $\varepsilon > 0$, there exists an $n_0(\varepsilon) \in \mathbb{N}$ such that, for any $n > n_0(\varepsilon)$, we can identify a typical set T_n with the properties that

1. $\mathbf{P}(\boldsymbol{u}^{(n)} \notin T_n) < \varepsilon$,
2. $\forall \boldsymbol{u}^{(n)} \in T_n$, we have $2^{-n(H+\varepsilon)} \leq \mathbf{P}(\mathcal{U}^{(n)} = \boldsymbol{u}^{(n)}) \leq 2^{-n(H-\varepsilon)}$.

Thus, for a source with the AEP, we will encode the typical strings with codewords of length $n(H+\varepsilon)$ and the atypical ones with an arbitrary encoding. We will obtain an effective encoding rate $H + o(1)$ bits per symbol.

We finish this section by particularizing this result for the binary case (when $U_i \in \mathcal{U}^{(n)}$ are distributed as a Bernoulli):

Theorem 2.3. For a Bernoulli source, the information rate equals the entropy of a source producing single letters; that is

$$H = h(U_j) = -\sum_{u\in I} p(u) \log_2 p(u).$$

Proof: For an IID sequence $U_1, U_2, \dots$ we have

$$p_n(\boldsymbol{u}^{(n)}) = \prod_{i=1}^{n} p(u_i) \Rightarrow -\log_2 p_n(\boldsymbol{u}^{(n)}) = \sum_{i=1}^{n} -\log_2 p(u_i).$$

Thus,

$$\xi_n = -\frac{1}{n}\log_2 p_n(\boldsymbol{u}^{(n)}) = \frac{1}{n}\sum_{i=1}^{n}\sigma_i,$$

where $\sigma_i = -\log_2 \mathbf{P}(U_i = u_i)$ are IID random variables. The expectation value of each σ_i is $\mathbb{E}\ \sigma_i = -\sum_j p(j)\log_2 p(j) = h$. So, $\mathbb{E}\ \xi_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\ \sigma_i = h$. The weak law of large numbers yields the convergence in probability $\xi_n \xrightarrow{\mathbf{P}} h$. □

2.3. Channel Coding

Formally, we characterize a channel by a conditional distribution

$$\mathbf{P}_{ch}(\boldsymbol{y}^{(N)} \text{ is received}|\boldsymbol{x}^{(N)} \text{ is sent}),$$

where $\boldsymbol{y}^{(N)} = y_1 \dots y_N$ is a word and $\boldsymbol{x}^{(N)} = x_1 \dots x_N$ is a codeword. We shall omit the super index N if it is understood from the context. We suppose this distribution is known to both sender and receiver, as well as the code (which converts the random text $\boldsymbol{u}^{(n)}$ into codewords $\boldsymbol{x}^{(N)}$ and is to recover $\boldsymbol{u}^{(n)}$ from $\boldsymbol{y}^{(N)}$. The goal of channel encoding is to successfully perform this recovery, despite the possible damage the information may have taken while traveling through the channel).

Definition 2.11. We say that a channel is memoryless if

$$\mathbf{P}_{ch}(\boldsymbol{y}^{(N)}|\boldsymbol{x}^{(N)}) = \prod_{i=1}^{N} P_{ch}(y_i|x_i).$$

Here, $P_{ch}(x|y)$ are the symbol-to-symbol channel probabilities. If the rows of the channel matrix P are permutations of each other, we say that the channel is symmetric.

Example 2.2. A memoryless binary symmetric channel where $x, y \in \{0, 1\}$ has a transition probability 2×2 matrix $(P_{ch}(y|x))_{xy}$ of the form

$$P_{ch} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

and p is called the symbol error probability or row error probability.

Definition 2.12. The two natural decoding rules are

(1) The ideal observer rule: Here, the receiver knows the probability distribution $\mathbf{P}_{src}$ of the source and hence the probabilities $p_N(\boldsymbol{x})$ of the codewords $\boldsymbol{x} \in \{0, 1\}^N$. We decode a received codeword $\boldsymbol{y}$ by a codeword $\boldsymbol{x}^*$ that maximizes the probability *a posteriori*:

$$\mathbf{P}(\boldsymbol{x} \text{ sent}|\boldsymbol{y} \text{ received}).$$

(2) The maximum likelihood rule: Here, the receiver does not know $p_N(\boldsymbol{x})$. We decode a received codeword $\boldsymbol{y}$ by a codeword $\boldsymbol{x}^*$ that maximizes the probability *a priori*:

$$\mathbf{P}(\boldsymbol{y} \text{ received}|\boldsymbol{x} \text{ sent}).$$

Definition 2.13. The rate R of a block encoding-decoding scheme is given by the ratio of the size of the message to the size of the corresponding codeword (both measured in bits):

$$R = \frac{\log \#\mathcal{M}}{N},$$

where $\mathcal{M}$ is the space of messages ($\#\mathcal{M} = 2^{NR}$).

R is equal to the number of bits of message transmitted per use of the channel. We say a rate R is achievable or reliable if there exists a sequence of encoding-decoding pairs ($f_N : \mathcal{M} \longrightarrow \{0,1\}^N, \hat{f}_N : \{0,1\}^N \longrightarrow \mathcal{M}$) such that the error probability $p(f_N, \hat{f}_N) := \max_{M \in \mathcal{M}} \mathbf{P}(\hat{f}_N(\boldsymbol{y}) \neq M | \boldsymbol{x} = f_N(M))$ is such that

$$p(f_N, \hat{f}_N) \longrightarrow 0 \quad \text{as} \quad N \longrightarrow \infty.$$

Finally, we define a channel's capacity:

Definition 2.14. The channel capacity is defined as

$$\mathcal{C} = \sup\{R : R \text{ is a reliable transmission rate}\}.$$

Theorem 2.4. (Shannon's second coding theorem)
For a memoryless channel, the capacity equals the maximum of the mutual entropy between a single input and a single output symbol. That is

$$\mathcal{C} = \sup_{p_X} \iota(X : Y),$$

where p_X stands for the distribution of the input symbol X and p_Y for the distribution of the output symbol Y:

$$p_Y(y) = \sum_{x'} p_X(x') \mathbf{P}_{ch}(y|x')$$

The proof of this theorem can be found in [**66**].

2.4. Current status of Coding Theory

In this last section we briefly discuss how can different kinds of codes approach Shannon's limit of channel capacity.

Definition 2.15. The channel rate is a measure in bits/s of the capacity [**54**]

$$C = W \cdot \log_2 \left(1 + \frac{S}{N}\right), \tag{2.1}$$

where W is a frequency in Hz called bandwidth of the channel, and S and N are the average power of the signal and the noise at the receiver, respectively.

The noise spectral-power density is the average energy of the noise per Hz, $N_0 = N/W$.

The quantity S/N is called SNR –signal to noise ratio– and is usually expressed in dB. This magnitude is often given via the energy per bit/power-spectral density $\rho = E_b/N_0$.

Both magnitudes are related via the spectral efficiency $\eta = C/W$:

$$\frac{S}{N} = \frac{C}{W} \cdot \frac{E_b}{N_0} = \eta \cdot \rho$$

We denote by p the bit error probability. In order to relate it with the S/N ratio [**54**], one should first consider the function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\tau^2/2} d\tau$: $p = Q(\sqrt{2E_b/N_0})$. The average energy per bit can also be found via $E_b = S/C$.

Remark 2.11. The channel rate (2.1) is closely related to the channel capacity (Definition 2.14), which is a number $\mathcal{C} \in [0, 1]$ that measures the maximum fraction of the information sent through the channel which is available at the receiving end, expressed in bits per transmission.

In Fig. 2.2, we have shown a plot of η vs. ρ and Shannon's limit, taken from (2.1), as the inequality for the suboptimal case, which can be rewritten as

$$C \leq W \log_2 \left(1 + \frac{S}{N}\right) \iff \rho \geq \frac{1}{\eta}(2^\eta - 1).$$

Above the line, no reliable encoding exists, according to Shannon's theorem. Under it, we have represented seven points [**71**] (the channel considered is binary symmetric):

- B corresponds to no encoding (9,6dB deficit, which is the horizontal distance from the point to the plotted line; note that the horizontal axis is plotted in logarithmic scale (dB), whereas the vertical axis, in linear scale),
- H corresponds to the Hamming $[7, 4, 3]$ code (7,8dB deficit),
- R corresponds to the repetition code $[3, 1, 3]$ (7,3dB deficit),
- G corresponds to the binary Golay code $[23, 12, 7]$ (5,8dB deficit),
- V corresponds to Voyager, 1986,
- G' corresponds to Galileo, 1989,
- T corresponds to a state of the art turbo-code

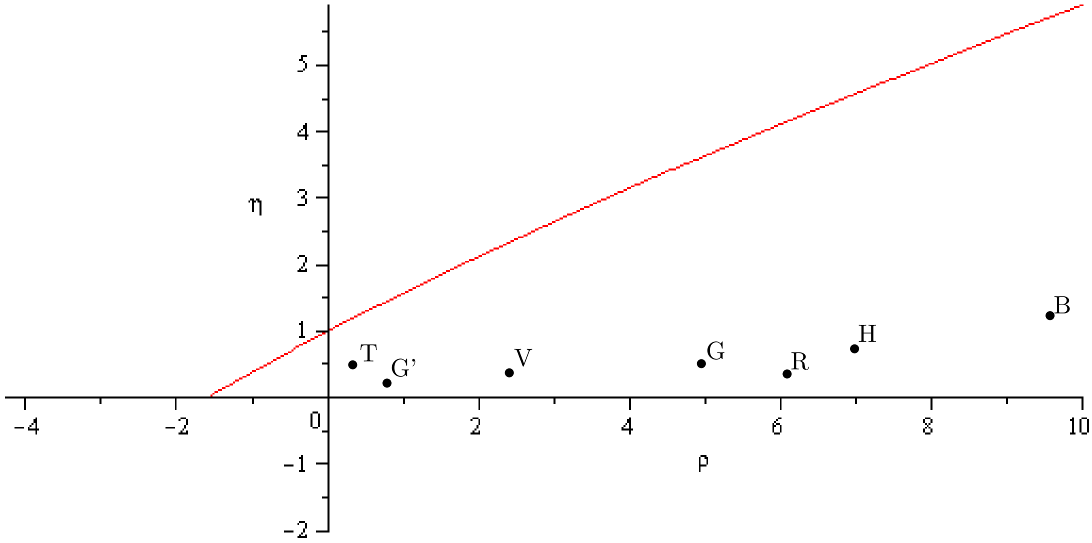

FIG. 2.2. Shannon's limit: $\rho > 10 \log_{10} \left(\frac{2^\eta - 1}{\eta}\right)$

Part 2

Physical Preliminaries

Chapter 3
Postulates of Quantum Mechanics

In this chapter we present the postulates of quantum mechanics and their mathematical description. We introduce the concept of qubit and its Bloch Sphere representation. We discuss the implications of performing measurements on a quantum system and the generalization to multi-qubit systems, as well as how a system evolves in time [**7, 8, 51, 46, 48**].

The mathematical model of von Neumann [**46**] is which has prevailed. To describe it, we begin this section with a *bottom-up* approach, in order to justify the axioms we state lately and interpret them in a *top-down* analysis.

3.1. Informal description

In classical mechanics, the state of a system of n particles at a time t_0 is completely determined by the positions $\{x_1(t_0), x_2(t_0), \ldots, x_n(t_0)\}$ and the velocities $\{\dot{x}_1(t_0), \dot{x}_2(t_0), \ldots, \dot{x}_n(t_0)\}$ of its particles at the time t_0. Given these initial conditions, Newton's laws of classical mechanics allow us to know the state of the system at any time t. Indeed, they are governed by a first-order differential equation in the variables $x_i, \dot{x}_i$ and, given a set of initial conditions, its solution is unique (Picard's theorem).
Quantum mechanics is based on a different framework:

To a physical quantum system Σ we associate a Hilbert space $\mathcal{H} = \mathcal{H}_\Sigma$; a state of Σ is completely described by a unit vector $|\psi\rangle$ (also called state vector or wave function) which resides in $\mathcal{H}$. The evolution in time of the state vector $|\psi\rangle$ is governed by the Schrödinger equation

$$i\hbar \frac{d}{dt}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle, \tag{3.1}$$

where $\hat{H}$ is a self-adjoint operator, also known as Hamiltonian of the system, and $\hbar = h/2\pi$, where h is Planck's constant[1].

The term self-adjoint is equivalent to Hermitian, so we will use any of them indistinctly.

Notation 3.1. It is customary in quantum mechanics the use of Dirac's notation, also known as bra-ket notation, in which the vectors of a Hilbert space are denoted by $|\psi\rangle$ (kets) and the corresponding covectors are denoted by $\langle\psi|$ (bras). Note that with this notation we are implicitly

[1] Planck's constant is a physical magnitude $h \approx 6,626 \cdot 10^{-34}$ Joule·second, experimentally determined.

identifying a Hilbert space with its dual. The main connection between a Hilbert space and its dual is given by the Riesz-Fréchet Representation Theorem. However, unless otherwise stated, we will work with finite-dimensional Hilbert spaces, in which we can do this identification. Note that the two spaces are then isomorphic, yet not canonically; for example, with a choice of basis we can explicit the identification.

Since Schrödinger equation is a linear differential equation of first order in time, the state $|\psi(t)\rangle$ is uniquely determined given the initial state $|\psi(t_0)\rangle$. Also, linearity implies that the superposition principle can be applied: $\alpha|\psi_1(t)\rangle + \beta|\psi_2(t)\rangle, \alpha, \beta \in \mathbb{C}$ is a solution if both $|\psi_i(t)\rangle$ are. Hence, the time-evolution operator $\hat{U}$, defined by

$$|\psi(t)\rangle = \hat{U}(t, t_0)|\psi(t_0)\rangle$$

is also linear.
When the Hamiltonian $\hat{H}$ is time independent, the solution to (3.1) reads

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}\hat{H}(t-t_0)}|\psi(t_0)\rangle,$$

where the exponential of the operator is defined as follows:

$$\hat{U}(t, t_0) = e^{-\frac{i}{\hbar}\hat{H}(t-t_0)} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{i}{\hbar}(t-t_0)\right)^n \hat{H}^n.$$

Remark 3.1. Given an Hermitian operator $\hat{A}$, $\hat{U} = e^{i\hat{A}}$ is unitary, with inverse $\hat{U}^\dagger = e^{-i\hat{A}}$. Conversely, any unitary operator $\hat{U}$ can be written as $\hat{U} = e^{i\hat{A}}$, where $\hat{A}$ is an Hermitian operator. This is known as Cayley's formula.

We associate with any observable A a self-adjoint operator $\hat{A}$, acting on the Hilbert space $\mathcal{H}_\Sigma$. The only possible outcome of a measurement of the observable A is one of the eigenvalues of the operator $\hat{A}$. By the spectral theorem, the eigenvalue equation for $\hat{A}$, $\hat{A}|i\rangle = a_i|i\rangle$, where $\{|i\rangle\}_i$ form an orthonormal basis of eigenvectors of A lets us expand the state vector $|\psi(t)\rangle = \sum_i c_i(t)|i\rangle$ over this basis. Then, the probability that a measurement of the observable A at time t results in outcome a_i is given by

$$p_i(t) = p(A = a_i|t) = |\langle i|\psi(t)\rangle|^2 = |c_i(t)|^2 . \tag{3.2}$$

Several remarks need to be made about this postulate:

Remark 3.2. Observables are the quantum analogue o dynamical variables in classical mechanics (e.g., position, linear and angular momentum, . . .). In contrast, other characteristics of the system (mass, electric charge) are not in the set of observables; they enter as parameters in the Hamiltonian of the system.

Remark 3.3. It could seem rather odd the association of physical observables with self-adjoint operators. This argument should help grasp its reason: The eigenvalues of a self-adjoint operator are real (just like the possible outcomes of a measurement) and its eigenvectors form a complete orthonormal set in the Hilbert space $\mathcal{H}_\Sigma$ associated with the system, each one uniquely determined up to a phase, if there is no degeneracy. Since we required $|\phi(t)\rangle$ to have unit norm, the probabilities are normalized:

$$\sum_i p_i(t) = \sum_i |c_i(t)|^2 = 1,$$

which means that the total probability of obtaining an outcome from the measurement of observable A is precisely 1.

So, an observable consists of giving a list of real values λ_i and a list of associated subspaces in which the corresponding value is satisfied. This is what determines an operator: a sum of projectors onto these subspaces, weighed by λ_i. The converse is given by Spectral Theorem (Theorem 1.4).

Example 3.1. Let $\mathcal{G} \subseteq \mathcal{H}$ be a linear subspace. The orthogonal projection $P_{\mathcal{G}} : \mathcal{H} \longrightarrow \mathcal{G}$ is an observable with eigenvalues 1 and 0, with associated eigenspaces $\mathcal{H}_1 = \mathcal{G}$ and $\mathcal{H}_0 = \mathcal{G}^{\perp}$.

Remark 3.4. If we consider the particular case $|\psi(t_0)\rangle = |i\rangle$, the measurement of A will yield the outcome a_i with probability 1. Thus, we will also call the eigenvectors of A eigenstates.

Remark 3.5. If $|\psi_1\rangle, |\psi_2\rangle$ are different normalized eigenvectors of A of respective eigenvalues a_1, a_2, the superposition principle tells us that $|\psi\rangle = \lambda_1|\psi_1\rangle + \lambda_2|\psi_2\rangle$ is also an allowed state for the system (with the normalization condition $|\lambda_1|^2 + |\lambda_2|^2 = 1$ and $\lambda_1, \lambda_2 \in \mathbb{C}$). Therefore, if we perform a measurement of A on a system described by the state $|\psi\rangle$, we will obtain, with probability $|\lambda_i|^2$, the outcome a_i.
This is *not equivalent* to a *naive* statistical mixture of the states $|\psi_1\rangle$ taken with probability $|\lambda_1|^2$ and $|\psi_2\rangle$ with probability $|\lambda_2|^2$.
We shall prove that a large number of N systems, all in the same state $|\psi\rangle$, is indeed not equivalent to an ensemble of $|\lambda_1|^2 N$ systems in the state $|\psi_1\rangle$ and $|\lambda_2|^2 N$ systems in the state $|\psi_2\rangle$:

Indeed, let us assume that we wish to calculate the probability $p(b_j)$ of obtaining outcome b_j for some observable B, with eigenvectors $|j\rangle$, in the system described by the state $|\psi\rangle$. This probability is

$$p(b_j) = |\langle j|\psi\rangle|^2 = |\lambda_1\langle j|\psi_1\rangle + \lambda_2\langle j|\psi_2\rangle|^2 ,$$

which can be expressed as

$$p(b_j) = |\lambda_1|^2 |\langle j|\psi_1\rangle|^2 + |\lambda_2|^2 |\langle j|\psi_2\rangle|^2 + 2\Re\{\lambda_1\lambda_2^*\langle j|\psi_1\rangle\langle j|\psi_2\rangle^*\}.$$

If we consider the statistical mixture described above, its result leads to

$$p_{mix}(b_j) = |\lambda_1|^2 |\langle j|\psi_1\rangle|^2 + |\lambda_2|^2 |\langle j|\psi_2\rangle|^2 .$$

The term $2\Re\{\lambda_1\lambda_2^*\langle j|\psi_1\rangle\langle j|\psi_2\rangle^*\}$ is called interference term.
With this example we have proved that the probability of obtaining b_j as the outcome of a measurement of B in a quantum-mechanical system, depends not only on the moduli $|\lambda_1|$ and $|\lambda_2|$, but on the relative phase between the complex numbers λ_1 and λ_2, which affects the product $\lambda_1\lambda_2^*$.

Remark 3.6. There may be cases in which we have degeneracies: Multiple eigenvectors $\{|n_s\rangle\}_{s=1\ldots g_n}$ for the same eigenvalue a_n. In this case it is convenient to introduce the projection operator

$$P_n = \sum_{s=1}^{g_n} |n_s\rangle\langle n_s|,$$

where $\langle n_s|$ is the dual of the vector $|n_s\rangle$ (Hermitian transposed in the matrix representation). When applied to a state, P_n operates as $P_n|\varphi\rangle = \left(\sum_{s=1}^{g_n} |n_s\rangle\langle n_s|\right)|\varphi\rangle = \sum_{s=1}^{g_n}(\langle n_s|\varphi\rangle)|n_s\rangle$. Thus, $|\varphi\rangle$ is projected onto the subspace generated by $\{|n_s\rangle\}_{s=1\ldots g_n}$ eigenvectors.

The postulate says in this case that the probability of outcome a_n is $\langle\psi|P_n|\psi\rangle$. Observe that for the non-degenerate case $g_s = 1$ and $\langle\psi|P_n|\psi\rangle = \langle\psi|n\rangle\langle n|\psi\rangle = \langle n|\psi\rangle\langle n|\psi\rangle^* = |\langle n|\psi\rangle|^2$ as we expected, recovering the rule from (3.2).

Remark 3.7. The density matrix.
In a practical situation, the state of a physical system is often not perfectly determined. For instance, a beam of atoms emitted by a thermal source: we do not know the kinetic energy of every atom, but only the distribution of their kinetic energies. We say that the information we have about the system is *incomplete.* We just know that the system is in a state taken from the ensemble

$$\{(p_1, |\psi_1\rangle), (p_2, |\psi_2\rangle), \ldots, (p_l, |\psi_l\rangle)\},$$

with the probabilities satisfying $\sum_i p_i = 1$. We say that we have a statistical mixture or a mixed state of the states $|\psi_k\rangle$, which are pure states. Observe that $|\psi_k\rangle$ need not to be orthogonal. The previous remark should suffice to convince ourselves that the statistical mixture of states $|\psi_k\rangle$ with weights p_k is not the same as the linear superposition

$$|\psi\rangle = \sum_k c_k|\psi_k\rangle, \quad |c_k|2 = p_k.$$

The question which arises is how do we describe such mixture? Can it be described by means of some 'average state vector'? The answer is no. However it is possible to describe it using an 'average operator': the density operator, as we shall see:

The probability $p(i)$ that a measurement of the observable A yields outcome a_i is given by

$$p(i) = \sum_{k=1}^{l} p_k\langle\psi_k|P_i|\psi_k\rangle,$$

where P_i is the projector onto the subspace associated with the eigenvalue a_i of A.

We can compute the mean value of the observable A, which we shall denote $\langle A\rangle$:

$$\langle A\rangle = \sum_{i=1}^{n} a_i p(i) = \sum_{k=1}^{l} p_k \sum_{i=1}^{n} a_i\langle\psi_k|P_i|\psi_k\rangle = \sum_{k=1}^{l} p_k\langle\psi_k|A|\psi_k\rangle.$$

Thus, we have probabilities appearing twice:

- In the initial lack of information on the system, characterized by weights p_k.
- In the measurement process, the probabilities $\langle\psi_k|P_i|\psi_k\rangle$ to obtain outcomes a_i from the measurement of observable A when the system is described by the state ψ_k. These probabilities are intrinsically quantum mechanical.

To take into account both, we introduce the density operator $\hat{\rho}$, defined as

$$\hat{\rho} \equiv \sum_k p_k|\psi_k\rangle\langle\psi_k|$$

Given a generic orthonormal basis $\{|i\rangle\}_{i=1\ldots n}$, where $n = \dim\mathcal{H}_\Sigma$, we can give $\hat{\rho}$ a matrix representation, known as the density matrix:

$$\rho = (\rho_{ij})_{ij} = (\langle i|\hat{\rho}|j\rangle)_{ij}.$$

The mean value of the operator A can also be computed by means of the density operator

$$Tr(\hat{\rho}A) = \sum_{i=1}^{n} \langle i|\rho A|i\rangle = \sum_{k=1}^{l}\sum_{i=1}^{n} p_k \langle i|\psi_k\rangle\langle\psi_k|A|i\rangle = \sum_{k=1}^{l}\sum_{i=1}^{n} p_k \langle\psi_k|A|i\rangle\langle i|\psi_k\rangle,$$

taking into account the completeness relation $\sum\limits_{i=1}^{n} |i\rangle\langle i| = \mathbb{I}$, we obtain

$$Tr(\hat{\rho}A) = \sum_{k=1}^{l} p_k \langle\psi_k|A|\psi_k\rangle = \langle A\rangle.$$

The mean value of A depends on the density operator and thus we will denote it $\langle A\rangle_\rho$. A similar argument proves that the probability $p(i)$ that a measurement of the observable A gives outcome a_i is equal to

$$p(i) = Tr(\hat{\rho}P_i).$$

We have showed that the density operator ρ completely characterizes the system, as we can predict the probabilities of the possible outcomes of any experiment performed on the system.

Properties 3.1. The density operator $\hat{\rho}$ has the following properties [8]:

- $\hat{\rho}$ is Hermitian.
- $\hat{\rho}$ has unit trace.
- $\hat{\rho}$ is a non-negative operator: For any vector $|\psi\rangle$ in $\mathcal{H}$ we have $\langle\psi|\hat{\rho}|\psi\rangle \geq 0$.

Remark 3.8. For a mixed state, one has $Tr\rho^2 < 1$, while for a pure state $Tr\rho^2 = 1$, enabling us to identify whether a state is pure or mixed given its density matrix.

Remark 3.9. Let us discuss the physical interpretation of the matrix elements of ρ:
Let us expand the pure state $|\psi_k\rangle = \sum\limits_{i=1}^{n} c_i^{(k)}|i\rangle$ over an orthonormal basis $\{|i\rangle\}$ of $\mathcal{H}$. The diagonal term

$$\rho_{ii} = Tr(\hat{\rho}P_i) = \sum_k p_k \left|c_i^{(k)}\right|^2, \quad P_i = |i\rangle\langle i|,$$

is the probability that the system is left in the state $|i\rangle$ after measuring the observable whose eigenstates are $\{|i\rangle\}$. Hence, we say that ρ_{ii} represents the population of the state $|i\rangle$.
The off-diagonal terms ρ_{ij}, called coherences, represent the interference between the states $|i\rangle$ and $|j\rangle$. This interference appears for any state $|\psi_k\rangle$ of the statistical mixture that contains a linear superposition of $|i\rangle$ and $|j\rangle$. Since

$$\rho_{ij} = \sum_k p_k c_i^{(k)} c_j^{(k)*},$$

ρ_{ij} may cancel despite individual terms may not. If $\rho_{ij} \neq 0$, it means that even after averaging over the statistical mixture, a quantum coherence effect will remain between the states $|i\rangle$ and $|j\rangle$.

Remark 3.10. The distinction between populations and coherences depends on the basis $\{|i\rangle\}$ we choose. Since ρ is Hermitian, non-negative and normalized (unit trace), via spectral decomposition, it is always possible to find an orthonormal basis $\{|m\rangle\}$ such that

$$\rho = \sum_m \alpha_m |m\rangle\langle m|, \quad 0 \leq \alpha_m \leq 1, \quad \sum_m \alpha_m = 1.$$

This means that the density matrix ρ can always be seen as a statistical mixture of the states $\{|m\rangle\}$, without coherences between them, although these states are not -in general- eigenstates of a physical observable.
This Schmidt decomposition is unique if the spectrum is non-degenerate (there is no multiple eigenvalue).

Remark 3.11. The density matrix corresponding to a pure state has rank 1 and in this case we can identify every unitary state vector with its corresponding density operator

$$|\psi\rangle \longleftrightarrow \hat{\rho} = |\psi\rangle\langle\psi|.$$

3.2. Axiomatic description

After this introduction, we can state the axioms in a more precise way:

Axiom 1. To any isolated quantum mechanical system Σ we associate a Hilbert space $\mathcal{H}$, called state space, and describe it completely with a unit vector in $\mathcal{H}$, the state vector[2].

The non-zero vectors $\boldsymbol{x} \in \mathcal{H}$ represent the pure states of Σ and two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$ represent the same state if, and only if, there exists $\xi \in \mathbb{C}$ such that $\boldsymbol{y} = \xi \boldsymbol{x}$.

In particular, $\boldsymbol{x}$ and the unit vector $\boldsymbol{x}/\left\|\boldsymbol{x}\right\|$ represent the same state. Moreover, two unit vectors represent the same state if and only if they differ in a unit factor (usually called a phaser). This means that the state space of Σ is $\mathbb{P}\mathcal{H}$, the projective space associated to $\mathcal{H}$.

Notation 3.2. Following Dirac's notation, we will write $|u\rangle$ the state corresponding to $[\boldsymbol{u}] \in \mathbb{P}\mathcal{H}$. Given two states $|u\rangle, |u\rangle' \in \mathcal{H}$ and two complex numbers $\lambda, \lambda' \in \mathbb{C}$, one can form the state $|\lambda \boldsymbol{u} + \lambda' \boldsymbol{u}'\rangle$ if $\lambda\lambda' \neq 0$. Such states are said to be in quantum superposition of the states $|u\rangle$ and $|u'\rangle$. By abuse of notation, such states are usually denoted as $\lambda|u\rangle + \lambda'|u'\rangle$.

Example 3.2. State of the spin (1/2).
A state of the spin of $\frac{1}{2}$ can be represented by the 2×2 matrix

$$\frac{1}{2}\begin{pmatrix} 1+x_3 & x_1 - ix_2 \\ x_1 + ix_2 & 1 - x_3 \end{pmatrix}, \quad x_i \in \mathbb{R}, \tag{3.3}$$

which is a density matrix if, and only if, $x_1^2 + x_2^2 + x_3^2 \leq 1$.

Axiom 2. The observables of a quantum mechanical system Σ are described by self-adjoint operators acting on the Hilbert space.

Example 3.3. Pauli matrices.
For a quantum spin 1/2 the matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

are called Pauli matrices and are used o describe the spin of directions x, y, z with respect to a coordinate system.

[2] We already showed the technical need of introducing density matrices to describe states, so Axiom 1 can be reformulated as *The physical states of a quantum mechanical system Σ are described by density operators acting on a Hilbert space $\mathcal{H}$.*

Any self-adjoint matrix is of the form $A_{x_0,\boldsymbol{x}} := x_0\sigma_0 + x_1\sigma_x + x_2\sigma_y + x_3\sigma_z$. Denoting the vectors $\boldsymbol{x} = (x_1, x_2, x_3), \boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ we can write the density matrix of the previous example as

$$\frac{1}{2}(\sigma_0 + \boldsymbol{x} \cdot \boldsymbol{\sigma}), \quad \|\boldsymbol{x}\| \leq 1.$$

This suggests a correspondence (a diffeomorphism) between 2×2 density matrices and the unit ball in $\mathbb{R}^3$. The extreme points of the ball correspond to pure states and any mixed state is the convex combination of pure states (with infinite possibilities).

3.3. The Qubit

The states of a spin-1/2 particle can be viewed as points lying on the unit sphere S^2. According to axiom 1, $\mathcal{H} = \mathbb{C}^2$ (spinor space).
The interpretation of this is as follows:

We begin by identifying a point $\xi = x + iy \in \mathbb{C}$ with the point $(x, y, 0) \in \mathbb{R}^3$.

In $\mathbb{R}^3$, we consider the manifold $S^2 = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 1\} \subset \mathbb{R}^3$ and we associate to ξ the point $P(\xi)$ obtained by means of the stereographic projection from $N = (0, 0, 1)$:

$$P(\xi) = \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}\right).$$

If we formally set $P(\infty) = N$, there is a bijection between S^2 and $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the completion of $\mathbb{C}$. More precisely, this bijection is given by $(x, y, z) \longmapsto \frac{x}{1-z} + i\frac{y}{1-z}$, for $z < 1$, and $N \longmapsto \infty$.

We can view $\overline{\mathbb{C}}$ as the projectivization of $\mathbb{C}^2$; this is the complex projective space of dimension 1:

$$\overline{\mathbb{C}} \cong \mathbb{P}(\mathbb{C}^2) = \mathbb{P}^1_{\mathbb{C}}.$$

Since any vector $(\xi_0, \xi_1) \in \mathbb{C}^2$ is proportional to a unique vector of the form $(1, \xi)$, where $\xi = \xi_1/\xi_0$ if $\xi_0 \neq 0$, and to $(0, 1)$ if $\xi_0 = 0$, we have the bijective map

$$\begin{array}{rcl}
\mathbb{P}^1_{\mathbb{C}} & \cong & \overline{\mathbb{C}} \\
{[\xi_0 : \xi_1]} & \longmapsto & \begin{cases} \xi = \xi_1/\xi_0 & \text{if} \quad \xi_0 \neq 0 \\ \infty & \text{if} \quad \xi_0 = 0 \end{cases} \\
{[1 : \xi]} & \longleftarrow & \xi \\
{[0 : 1]} & \longleftarrow & \infty
\end{array}$$

This is what justifies that we take $\mathcal{H} = \mathbb{C}^2$.

Definition 3.1. The sphere S^2, with the structure of $\mathbb{P}^1_{\mathbb{C}}$, is called the Riemann sphere, which is the simplest compact Riemann surface.
However, in quantum contexts, it is often called the (Poincaré) Bloch sphere.

Remark 3.12. It is natural to analyze the discussion above in spherical coordinates. Let $P = (x, y, z) \in S^2 \subset \mathbb{R}^3$, $\varphi = \arg(x + iy)$ and θ be the angle between OP and ON, where O is the center of the sphere $(0, 0, 0)$.

Given the change of coordinates from spherical to Cartesian coordinates

$$x = \sin\theta\cos\varphi, \quad y = \sin\theta\sin\varphi, \quad z = \cos\theta,$$

one obtains the following expression for $P(x,y,z) \in \overline{\mathbb{C}}$:

$$\xi = \frac{x}{1-z} + i\frac{y}{1-z} = \frac{\sin\theta\cos\varphi}{1-\cos\theta} + i\frac{\sin\theta\sin\varphi}{1-\cos\theta} = \frac{\sin\theta}{1-\cos\theta}e^{i\varphi}$$

This can be written in the more compact expression

$$\xi = e^{i\varphi}\cot\frac{\theta}{2}.$$

In $\mathbb{P}^1_{\mathbb{C}}$, this corresponds to the point

$$[1 : e^{i\varphi}\cot\frac{\theta}{2}] \sim [e^{-i\varphi/2}\sin\frac{\theta}{2} : e^{i\varphi/2}\cos\frac{\theta}{2}] \in \mathbb{P}^1_{\mathbb{C}} ;$$

thus, we shall write in the chosen notation

$$P \longleftrightarrow |p\rangle = e^{-i\varphi/2}\sin\frac{\theta}{2}|0\rangle + e^{i\varphi/2}\cos\frac{\theta}{2}|1\rangle \in \mathbb{P}^1_{\mathbb{C}} \cong S^2. \tag{3.4}$$

In Fig. 3.1 we can see the representation of the Bloch sphere[3].

Remark 3.13. Equation (3.4) allows for the following geometric interpretation of unitary matrices U and rotations of S^2: A unitary matrix $U = \begin{pmatrix} u_0 & u_1 \\ -u_1^* & u_0^* \end{pmatrix}$ can be viewed as a linear map $\mathbb{C}^{\,2} \longrightarrow \mathbb{C}^{\,2}$ such that the row vector $\boldsymbol{\xi} = [\xi_0, \xi_1]$ is transformed as $\boldsymbol{\xi} = \boldsymbol{\xi}U^T$.

In $\mathbb{P}^1_{\mathbb{C}}$, it induces a projective map given by

$$[\xi_0, \xi_1] \longmapsto [u_0\xi_0 + u_1\xi_1 : -u_1^*\xi_0 + u_0^*\xi_1],$$

Hence, we have a map $\overline{U} : \overline{\mathbb{C}} \longrightarrow \overline{\mathbb{C}}$ defined by

$$\xi \mapsto \frac{u_0^*\xi - u_1^*}{u_1\xi + u_0}, \qquad \infty \mapsto u_0^*/u_1,$$

where $\xi = \xi_1/\xi_0$. In turn, this corresponds to the map $\widetilde{U} : S^2 \longrightarrow S^2$ such that $\widetilde{U}(P(\xi)) = P(\overline{U}(\xi))$.

For instance, if U^T is one of the following matrices,

$$R_z(\alpha) = \begin{pmatrix} e^{-i\frac{\alpha}{2}} & 0 \\ 0 & e^{i\frac{\alpha}{2}} \end{pmatrix}, \quad R_y(\beta) = \begin{pmatrix} \cos(\frac{\beta}{2}) & -\sin(\frac{\beta}{2}) \\ \sin(\frac{\beta}{2}) & \cos(\frac{\beta}{2}) \end{pmatrix}, \quad R_x(\gamma) = \begin{pmatrix} \cos(\frac{\gamma}{2}) & i\sin(\frac{\gamma}{2}) \\ i\sin(\frac{\gamma}{2}) & \cos(\frac{\gamma}{2}) \end{pmatrix}$$

Then, $\widetilde{R_z(\alpha)}, \widetilde{R_y(\beta)}, \widetilde{R_x(\gamma)}$ are the rotations of angles α, β, γ around axis z, y, x, respectively (positively-oriented, or following the right-hand rule).

[3] Any physical quantum system which can take two states can be a qubit. For example, we can choose a single photon as our qubit and use its polarization as a degree of freedom to represent its states: If we say that $|0\rangle$ means *horizontally* polarized and $|1\rangle$ *vertically* polarized (we have chosen a basis), then the points $|0\rangle + i|1\rangle$, $|0\rangle - |1\rangle$, $|0\rangle - i|1\rangle$ and $|0\rangle + |1\rangle$ would be *right circular* polarization, *45º linear* polarization, *left circular* polarization and *-45º linear* polarization, respectively.

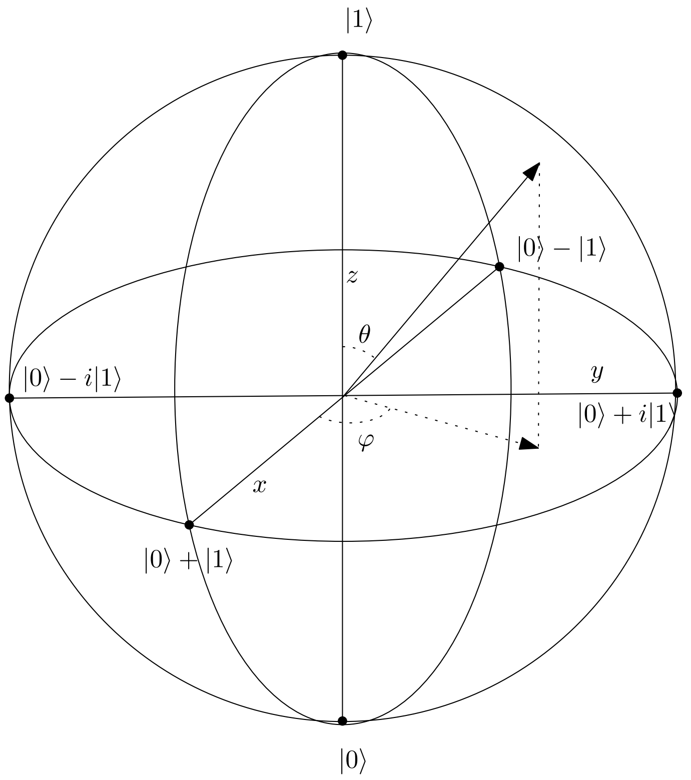

FIG. 3.1. Bloch sphere for the representation of a qubit.

Finally, we recover Equation (3.4) from the transformation

$$|1\rangle \xrightarrow{R_z(\varphi)R_y(\theta)} |p\rangle = e^{-i\varphi/2}\sin\frac{\theta}{2}|0\rangle + e^{i\varphi/2}\cos\frac{\theta}{2}|1\rangle,$$

which, in S^2, corresponds to

$$P = \widetilde{R_z(\varphi)}\widetilde{R_y(\theta)}N.$$

3.4. Measurements

Axiom 3. Let $\mathcal{X}$ be a finite set and for $x \in \mathcal{X}$ an operator $V_x \in \mathcal{B}(\mathcal{H})$ such that

$$\sum_{x\in\mathcal{X}} V_x^\dagger V_x = \mathbb{I}.$$

This sum is also called a resolution of the identity. Such an indexed family of operators is a model of a measurement with values in $\mathcal{X}$. If the measurement is performed in a state ρ, then the outcome $x \in \mathcal{X}$ appears with probability $Tr(V_x\rho V_x^\dagger)$ and, after the measurement, the state of the system is

$$\frac{V_x\rho V_x^\dagger}{Tr(V_x\rho V_x^\dagger)}$$

Remark 3.14. As a particular case, one can consider the measurement of an observable described by a self-adjoint operator A with spectral decomposition $\sum_i \lambda_i E_i$. In this case, $\mathcal{X} = \{\lambda_i\}$ is the set of eigenvalues and $V_i = E_i$. We have already shown that the expectation of the random outcome is $Tr(\rho A)$.

In the case of pure state vectors, if the wave function of the system, which immediately before the measurement of observable A was in the state $|\psi\rangle$, immediately after the measurement collapses onto the state $P_n|\psi\rangle/\sqrt{\langle\psi|P_n|\psi\rangle}$, where P_n is the projection operator over the subspace corresponding to a_n. All subsequent measurements of A will lead to the same result with probability 1 as the state has already collapsed [**46, 47**].

3.5. Composite systems

Axiom 4. The composite system of subsystems with associated Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ is described by the tensor product Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$.

When $A_i \in \mathcal{B}(\mathcal{H}_i)$, the action of the tensor product operator $A_1 \otimes A_2$ is determined by

$$(A_1 \otimes A_2)(|\eta_1\rangle \otimes |\eta_2\rangle) = A_1|\eta_1\rangle \otimes A_2|\eta_2\rangle.$$

When $A = A^\dagger$ is an observable of the first system, its expectation value in the vector state $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is

$$\langle\psi|A \otimes \mathbb{I}_2|\psi\rangle,$$

where $\mathbb{I}_2$ is the identity operator acting on $\mathcal{H}_2$.
So, when we want to consider an observable acting on the total system, extending it by tensorizing with identity operators on the rest of subsystems will suffice.

Notation 3.3. It is common the abuse of notation $|e\rangle \otimes |f\rangle \equiv |e\rangle|f\rangle \equiv |e, f\rangle \equiv |ef\rangle$.

Lemma 3.1. If $\mathcal{H}_1, \mathcal{H}_2$ are finite dimensional Hilbert spaces, with respective basis $\{|e_j\rangle\}$ and $\{|f_i\rangle\}$. Let

$$|\varphi\rangle = \sum_{i,j} w_{ij}|e_j\rangle \otimes |f_i\rangle$$

be the expansion of a unit vector $|\varphi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ and W the matrix $(w_{kl})_{kl}$. Then [**51**] $W^\dagger W$ is a density matrix and

$$\langle\varphi|(A \otimes \mathbb{I}_2)|\varphi\rangle = Tr(AW^\dagger W).$$

This lemma shows a natural way from state vectors to density matrices. Given a density matrix ρ in $\mathcal{H}_1 \otimes \mathcal{H}_2$, there exist density matrices $\rho_i \in \mathcal{B}(\mathcal{H}_i)$ such that

$$\langle A \otimes \mathbb{I}_2\rangle_\rho = Tr((A \otimes \mathbb{I}_2)\rho) = Tr(A\rho_1) = \langle A\rangle_{\rho_1}$$

and

$$\langle \mathbb{I}_1 \otimes B\rangle_\rho = Tr((\mathbb{I}_1 \otimes B)\rho) = Tr(B\rho_2) = \langle B\rangle_{\rho_2}.$$

The matrices ρ_1 and ρ_2 are called reduced density matrices and they are the quantum analogue of marginal distributions. We also say that we have obtained the partial trace of one of the subsystems or that we have 'traced out' the rest of the subsystems. More precisely:

$$\begin{array}{rlcc} Tr_2: & \mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_2) & \longrightarrow & \mathcal{B}(\mathcal{H}_1) \\ & A \otimes B & \longmapsto & A \cdot Tr(B) \end{array}$$

$$\begin{array}{rlcr} Tr_1: & \mathcal{B}(\mathcal{H}_1)\otimes\mathcal{B}(\mathcal{H}_2) & \longrightarrow & \mathcal{B}(\mathcal{H}_2) \\ & A\otimes B & \longmapsto & B\cdot Tr(A) \end{array}$$

so that we have

$$\rho_1 = Tr_2(\rho), \quad \rho_2 = Tr_1(\rho)$$

3.6. State transformations

Axiom 5. If Σ lies in a non-reactive environment (i.e., the environment is not affected by Σ) in the interval $[0,t]$, there exists a unitary operator $\hat{U}:\mathcal{H}\longrightarrow\mathcal{H}$ such that $|u_t\rangle = \hat{U}|u_0\rangle$.

Example 3.4. Conservative systems.

When the Hamiltonian H of a system does not depend explicitly on time, we say that the system is conservative. It is a result from classical mechanics that the energy E of the system is constant in time (it is a constant of motion). In quantum mechanics context, the solution to Schrödinger equation (3.1) can be written easily, once we know the eigenvalues E_n and the eigenvectors $|n\rangle$ of the Hamiltonian operator H: $H|n\rangle = E_n|n\rangle$; and they are time-independent too. The eigenstates are called stationary states and if a system Σ is described by such states, its physical properties do not change in time.

3.7. Entanglement: an introductory description

Let us consider the Hilbert space of two spins 1/2, which is $\mathbb{C}^{\,2}\otimes\mathbb{C}^{\,2}$. In this space the vectors

$$|e_1\rangle := |\uparrow\rangle\otimes|\uparrow\rangle, \quad |e_2\rangle := |\uparrow\rangle\otimes|\downarrow\rangle, \quad |e_3\rangle := |\downarrow\rangle\otimes|\uparrow\rangle, \quad |e_4\rangle := |\downarrow\rangle\otimes|\downarrow\rangle$$

form a basis. The vector state (known as EPR pair)

$$|\phi\rangle = \frac{|\uparrow\rangle\otimes|\downarrow\rangle - |\downarrow\rangle\otimes|\uparrow\rangle}{\sqrt{2}} \tag{3.5}$$

has the following property:
If we consider the observable

$$A := \sum_{i=1}^{4} i|e_i\rangle\langle e_i|,$$

which has eigenvalues $1,2,3,4$ and the basis vectors are its eigenvectors.
When we perform a measure of the observable A on the system prepared in the state $|\phi\rangle$ we obtain the results $1,2,3,4$ with respective probabilities $0,\frac{1}{2},\frac{1}{2},0$. Therefore, in the vector state $|\varphi\rangle$ the spins are anti-correlated. This[4] will be explained more precisely in Section 4.1.1.

We will treat this phenomenon, known as entanglement, in following chapters.

[4] If we perform a measurement on Hilbert space 1 and, say, it collapses to state $|\uparrow\rangle$, the EPR pair will have collapsed onto the state $|\uparrow\rangle\otimes|\downarrow\rangle$, so when performing a measurement on Hilbert space 2, we will measure $|\downarrow\rangle$ with probability 1. Similarly, if we measure Hilbert space 2 and the EPR pair collapses onto $|\uparrow\rangle$, then $|\phi\rangle$ will have collapsed to state $|\downarrow\rangle\otimes|\uparrow\rangle$ (global phase is irrelevant) and in Hilbert space 1 we will be measuring state $|\downarrow\rangle$ with probability 1.
One could ask what happens if the measurements are performed simultaneously, or in causally disconnected environments. This paradox will be solved in the next section, with Theorem 4.2.

Chapter 4
Quantum Probability

In this chapter is described how quantum probability is essentially different from classical probability (Kolmogorov's axioms). We show how Hidden Variable Theories do not differ from quantum mechanics via the CHSH inequality and a counterexample based on Aspect's experiment. We also introduce the need of *-algebras of operators and states to devise a non-commutative probability theory [**40**].

We discuss the quantum impossibilities which arise when making probability non-commutative (no cloning theorem, no-classical-coding theorem, or that information cannot travel faster than light) as well as several quantum novelties which appear, such as quantum teleportation or quantum dense coding.

4.1. Bell inequalities

Perhaps the most spectacular, counter-intuitive manifestation of quantum mechanics is the phenomenon known as entanglement, which can be observed in composite quantum systems. Let us discuss the problem. In the simplest case of a bipartite quantum system, we have

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$$

Definition 4.1. A pure state $|\psi\rangle \in \mathcal{H}$ is said to be entangled or non-separable if it cannot be written as a simple tensor product of a state $|\alpha\rangle_1 \in \mathcal{H}_1$ and $|\beta\rangle_2 \in \mathcal{H}_2$. Otherwise, if we can write

$$|\psi\rangle = |\alpha\rangle_1 \otimes |\beta\rangle_2$$

the state $|\psi\rangle$ is separable (composite, or product state are names which can also be found in the literature).

When two systems are entangled, we cannot assign them individual state vectors $|\alpha\rangle_1$ and $|\beta\rangle_2$. The intriguing non-classical properties of entangled states were illustrated by Einstein, Podolsky and Rosen [**23**] in 1935, showing that quantum theory leads to a contradiction with the two -apparently natural- assumptions below:

(i) *Reality principle*: If we can predict with certainty the value of a physical quantity, then this value has physical reality, independently of our observation. E.g., if a system's wave function $|\psi\rangle$ is an eigenstate of an operator A: $A|\psi\rangle = a|\psi\rangle$, then the value a of the observable A is an element of physical reality, as it will be the outcome of the measurement of A with probability 1.

(ii) *Locality principle*: If two systems are causally disconnected, the result of any measurement on one system cannot influence the result of a measurement performed on the second system. According to relativity theory, two events are causally disconnected if $(\Delta x)^2 > c^2(\Delta t)^2$, where the inequality terms are separation in space, speed of light and separation in time, respectively, in some inertial reference frame.

In quantum mechanics, if two operators $\hat{A}, \hat{B}$ do not commute, then the two physical quantities corresponding to the operators $\hat{A}$ and $\hat{B}$ cannot have simultaneous reality (we cannot predict with certainty the outcome of a simultaneous measurement of both $\hat{A}$ and $\hat{B}$), since due to Heisenberg's principle, a measurement of $\hat{A}$ destroys knowledge of $\hat{B}$.

4.1.1. The CHSH inequality. A natural question arises: if we cannot predict with certainty the outcome of measurements, maybe quantum mechanics is not complete, in the sense that there may exist a hidden variable λ such that there is a well-defined characterization, in terms of λ, of the probabilities associated to $O(\lambda)$: the result obtained from the measurement of the physical observable O, for all λ. This framework is known as "Hidden Variable Theory". In this section we will show that, even with this assumption, local realism leads to a contradiction:

The distribution probability $\rho(\lambda)$ of the variable λ needs to be such that the average values predicted by quantum mechanics are recovered:

$$\langle O \rangle = \int O(\lambda)\rho(\lambda)d\lambda.$$

Now let us consider the EPR *gedanken* experiment:

We have a source S that emits pairs of spin-1/2 particles between Alice (A) and Bob (B) in the entangled EPR state:

$$|\phi\rangle = \frac{|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

The first particle is sent to Alice and the second is sent to Bob. According to Quantum mechanics, if Alice measures the z component of the spin of her particle and obtains, let us say, $\sigma_z^{(A)} = +1$ then $|\phi\rangle$ will collapse to $|01\rangle$ ($|0\rangle$ is the eigenstate of σ_z with $+1$ eigenvalue and $|1\rangle$ corresponds to -1) and Bob will measure $\sigma_z^{(B)} = -1$ with probability 1. However, $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ are the eigenstates of σ_x with respective eigenvalues $+1, -1$ and we can also write $|\phi\rangle = \frac{1}{\sqrt{2}}(|+-\rangle - |-+\rangle)$; this means that if Alice measures, for instance, $\sigma_x^{(A)} = +1$, then $|\phi\rangle$ will collapse to $|+-\rangle$ and the state of the Bob's particle collapses onto an eigenstate of $\sigma_x^{(B)}$. This means that we cannot simultaneously assign an element of physical reality with both observables, and the reason is that they don't have the same eigenstates; hence, they don't commute: $[\sigma_x^{(B)}, \sigma_z^{(B)}] \neq 0$. Since Alice can choose which observable to measure, even after the particles have separated an arbitrary distance, this contradicts the locality principle: she should not be able to modify Bob's particle. If we want to accept locality, then realism has to be dropped: the wave function is not seen as a physical object, but just a mathematical tool to predict probabilities for the outcome of experiments.

Let's now introduce a hidden variable to our model: We call $A(\boldsymbol{a}, \lambda)$ and $B(\boldsymbol{b}, \lambda)$ the results of the measurements of the spin polarizations $\boldsymbol{\sigma}^{(A)} \cdot \boldsymbol{a}$ and $\boldsymbol{\sigma}^{(B)} \cdot \boldsymbol{b}$ along the directions $\boldsymbol{a}$ and $\boldsymbol{b}$, performed by Alice and Bob, respectively.

According to the locality principle, the outcome of Alice's measurements cannot depend on the outcome of Bob's measurements. This means that the mean value of the correlations between their polarization measurements is given by

$$C(\boldsymbol{a}, \boldsymbol{b}) = \int A(\boldsymbol{a}, \lambda) B(\boldsymbol{b}, \lambda) \rho(\lambda) d\lambda.$$

For instance, quantum mechanics predicts perfect anticorrelation for the Bell state when $\boldsymbol{a} = \boldsymbol{b}$ and therefore $C(\boldsymbol{a}, \boldsymbol{a})_{quantum} = -1$.

We will compute $C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'})$ and bound this quantity:

$$C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'}) = \int (A(\boldsymbol{a}, \lambda) B(\boldsymbol{b}, \lambda) - A(\boldsymbol{a}, \lambda) B(\boldsymbol{b'}, \lambda)) \rho(\lambda) d\lambda =$$

$$\int A(\boldsymbol{a}, \lambda) B(\boldsymbol{b}, \lambda)(1 \pm A(\boldsymbol{a'}, \lambda) B(\boldsymbol{b'}, \lambda)) \rho(\lambda) d\lambda - \int A(\boldsymbol{a}, \lambda) B(\boldsymbol{b'}, \lambda)(1 \pm A(\boldsymbol{a'}, \lambda) B(\boldsymbol{b}, \lambda)) \rho(\lambda) d\lambda$$

Observe that $|A(\boldsymbol{a}, \lambda)| = |B(\boldsymbol{b}, \lambda)| = 1$, since they are polarization measurements; also, $\rho(\lambda) \geq 0 \forall \lambda$. Thus,

$$|C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'})| \leq \int (1 \pm A(\boldsymbol{a'}, \lambda) B(\boldsymbol{b'}, \lambda)) \rho(\lambda) d\lambda + \int (1 \pm A(\boldsymbol{a'}, \lambda) B(\boldsymbol{b}, \lambda)) \rho(\lambda) d\lambda,$$

which can be rewritten as

$$|C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'})| \leq \pm (C(\boldsymbol{a'}, \boldsymbol{b'}) + C(\boldsymbol{a'}, \boldsymbol{b})) + 2 \int \rho(\lambda) d\lambda.$$

Therefore,

$$|C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'})| \leq -|C(\boldsymbol{a'}, \boldsymbol{b'}) + C(\boldsymbol{a'}, \boldsymbol{b})| + 2 \int \rho(\lambda) d\lambda,$$

Using the normalization $\int \rho(\lambda) d\lambda = 1$ we arrive at the Clauser, Horne, Shimony and Holt (CHSH) inequality [**16**]:

$$|C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'})| + |C(\boldsymbol{a'}, \boldsymbol{b'}) + C(\boldsymbol{a'}, \boldsymbol{b})| \leq 2. \tag{4.1}$$

Remark 4.1. The main point is that there exists a set of directions $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a'}, \boldsymbol{b'})$ such that, considering entangled states, quantum mechanics violates the CHSH inequality.

An example of such set of directions is represented in Fig. 4.1.

Example 4.1. For the set of directions shown in Fig. 4.1, and the EPR state $|\phi\rangle$, quantum mechanics predicts that $C(\boldsymbol{a}, \boldsymbol{b}) = -\boldsymbol{a} \cdot \boldsymbol{b} = -\cos(\theta_{ab})$, being θ_{ab} the angle between the directions $\boldsymbol{a}$ and $\boldsymbol{b}$; thus we have

$$\{|C(\boldsymbol{a}, \boldsymbol{b}) - C(\boldsymbol{a}, \boldsymbol{b'})| + |C(\boldsymbol{a'}, \boldsymbol{b'}) + C(\boldsymbol{a'}, \boldsymbol{b})|\}_{quantum} =$$

$$|-\cos\theta + \cos 3\theta| + |-\cos\theta - \cos\theta| = 2\sqrt{2} \not\leq 2$$

when $\theta = \pi/4$.

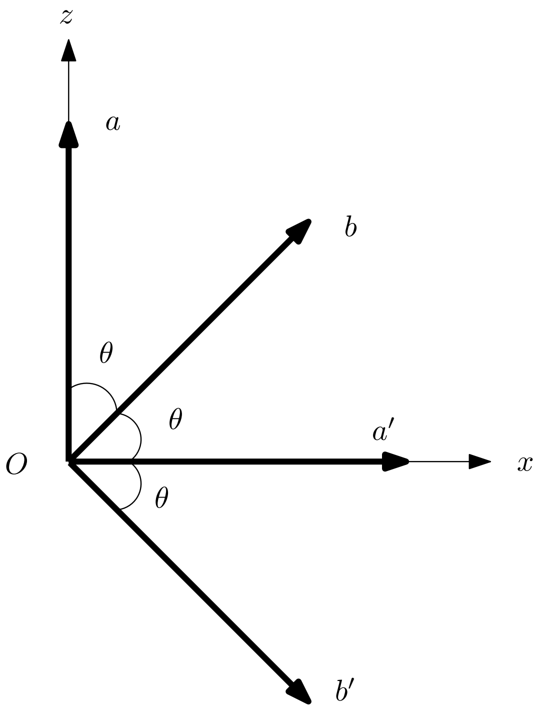

Fig. 4.1. Choice of directions that lead to a violation of the CHSH inequality with $\theta = \pi/4$. With two polarizers oriented according to a and b we obtain the correlation function $C(a, b)$.

4.2. Classical-quantum probability

In classical probability, according to Kolmogorov's axioms ($\sim$ 1930) a probability space $(\Omega, \sigma, \mathbf{P})$ is determined by giving a set Ω of outcomes ω, a σ-algebra σ which specifies which subsets $S \subset \Omega$, $S \in \sigma$, which are to be considered as events, and by associating a probability function $\mathbf{P}(S) \in [0, 1]$ to each one of these events.

Remark 4.2. The $\sigma-$algebra of events σ is a collection of sets closed with respect to all possible (countable) unions and intersections of its elements, which are called measurable.
The probability measure $\mathbf{P}$ must be σ-additive; namely, the probability of any union $S = \bigcup_j S_j$ of disjoint measurable subsets $S_j \cap S_k = \emptyset$, must be the sum of the probabilities of the corresponding subsets, $\mathbf{P}(S) = \sum_j \mathbf{P}(S_j)$ and normalized $\mathbf{P}(\Omega) = 1$.

Remark 4.3. This is a special case of measure theory, in which the measure $\mathbf{P}$ is not restricted to be between 0 and 1.

In quantum probability, we will weaken this scheme. We abandon the notion that a point $\omega \in \Omega$ decides about the occurrence or non-occurrence of all events in a simultaneous way. It should be natural by now if we take as events certain closed subspaces of a Hilbert Space or, equivalently, a set of projections to which we will associate probabilities.

More precisely,

(1) The set $\mathcal{E}$ of all events of a quantum model must be the set of projections in some $*$-algebra $\mathcal{A}$ of operators on $\mathcal{H}$.
(2) The probability function $\mathbf{P} : \mathcal{E} \longrightarrow [0,1]$ must be σ-additive.

4.3. *-algebras of operators and states

Definition 4.2. By a (unital) *-algebra of operators on $\mathcal{H}$ we mean a subspace $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ of the space of all linear maps $A : \mathcal{H} \longrightarrow \mathcal{H}$ such that $\mathbb{I} \in \mathcal{A}$ and

$$A, B \in \mathcal{A} \Rightarrow \lambda A, A + B, A \cdot B, A^{\dagger} \in \mathcal{A}.$$

A state on $\mathcal{A}$ is a linear functional $\varphi : \mathcal{A} \longrightarrow \mathbb{C}$ satisfying

(1) $\forall A \in \mathcal{A}, \quad \varphi(A^{\dagger}A) \geq 0$,
(2) $\varphi(\mathbb{I}) = 1$.

A pair $(\mathcal{A}, \varphi)$ is a quantum probability space.

Example 4.2. Let $P_1, P_2, \ldots, P_k$ be mutually orthogonal projections on $\mathcal{H}$, with sum $\mathbb{I}$.

Their linear span

$$\mathcal{A} = \left\{ \sum_{j=1}^{k} \lambda_j P_j \mid \lambda_j \in \mathbb{C} \right\}$$

forms a unital *-algebra of operators on $\mathcal{H}$. If $\boldsymbol{\psi}$ is some unit vector in $\mathcal{H}$, it determines a state φ by $\varphi(A) := \langle \boldsymbol{\psi}, A\boldsymbol{\psi} \rangle_{\mathcal{H}}$. The probabilities of this model are $p_j := \varphi(P_j) = \|P_j\boldsymbol{\psi}\|^2$, and they correspond to classical probabilities, since mutually orthogonal projections commute.

Example 4.3. Let $\mathcal{A}$ be the *-algebra of all complex $n \times n$ matrices, $\mathcal{A} = M_n(\mathbb{C}\ )$. Let $A \in \mathcal{A}$ and $\varphi(A) := Tr(\rho A)$, where $\rho \geq 0$ and $Tr(\rho) = 1$.

A is thought as an observable if $A = A^{\dagger}$ and the expected value of the observable A, given that the system is in the state φ, is $\varphi(A)$. This corresponds to the purely quantum mechanical situation.

As particular cases, $n = 2$ corresponds to the qubit described in Section 3.3. If the state is pure, that is, $\rho = |\psi\rangle\langle\psi|$, for some unit vector $|\psi\rangle \in \mathcal{H}$, then the expected value of the observable A is $\varphi(A) = Tr(\rho A) = Tr(|\psi\rangle\langle\psi|A) = Tr(\langle\psi|A|\psi\rangle) = \langle\psi|A|\psi\rangle$; moreover, if all the matrices in $\mathcal{A}$ commute, we are in the previous example situation.

4.4. Quantum impossibilities

Theorem 4.1. (No-cloning)
Let $|s\rangle \in \mathcal{H}$. There does not exist a unitary evolution U such that

$$|\psi\rangle \otimes |s\rangle \overset{U}{\longmapsto} U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle$$

for all $|\psi\rangle \in \mathcal{H}$.

Proof: If this procedure works for two particular pure states, $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$, then we have

$$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle$$

and

$$U(|\varphi\rangle \otimes |s\rangle) = |\varphi\rangle \otimes |\varphi\rangle.$$

Since U is unitary and $|s\rangle$ is a pure state, taking the scalar product of the previous equations implies

$$\langle\psi|\varphi\rangle = (\langle\psi|\varphi\rangle)^2,$$

which means that, either $|\psi\rangle = |\varphi\rangle$ or $|\psi\rangle$ and $|\varphi\rangle$ are orthogonal states. □

Remark 4.4. If a quantum device were to clone quantum states, it could only clone states orthogonal to one another. For example, it could not copy $|0\rangle$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ since they are not orthogonal states.

Theorem 4.2. (No faster than light communication)
Information cannot travel faster than light's speed c.

The non-locality of quantum entangled states does not allow any transmission of information faster than light; thus, it does not conflict with special relativity [**23, 26, 48**].

Before discussing the general theorem, let us discuss a simple example.

Example 4.4. Let an entangled pair be shared between two observers, Alice and Bob. Alice performs on her particle a spin test along the z-direction. She obtains perfectly random outcomes in $\{-1, 1\}$, each with the same probability, forming a sequence. If Bob (sufficiently separated) performs the same test, he will obtain perfectly correlated results. However, his results are also perfectly random and Bob has no way to know if Alice performed a measurement unless they use classical communication.
If Bob were able to clone each particle he receives, the situation changes: Creating $4N$ copies of Bob's particle, suppose Alice chooses to perform a test along the z-direction or along a direction at 45° with respect to z. Bob then sends N copies to 4 different apparatuses, measuring in the directions $z, -z$ or the two conjugate of the second basis.
Let us assume the entangled pair corresponds to the Bell state described by (3.5). When Alice gets outcome 1 (-1) in the z basis, Bob's results will be N (0) giving outcome -1 and 0 (N) giving outcome -1, in the same basis. In the other basis, the results will be completely random (since the state and the measurement basis are not orthonormal) and he will expect to get $N/2$ particles per result. When Alice measures in the second basis, an analogous result applies for Bob.
Therefore, by simply inspecting which detector counts 0 events, Bob can know which basis Alice used to measure her state, a knowledge that could be easily used to implement faster than light communication.
Nevertheless, Theorem 4.1 does not allow for such scheme.

Let us now sketch the proof of Theorem 4.2:

Suppose we perform a POVM on the subsystem Σ_1 of a composed system $\Sigma_1 + \Sigma_2$ described by the statistical operator ρ_{12}. This corresponds to projecting onto an eigenstate $|s\rangle$, an operation which we can describe with the projection operator $P_s^1 = |s\rangle\langle s|$.

Alice performs a measurement on Σ_1; thus, the state becomes now

$$\rho_{12} = \rho'_{12} = \sum_s P_s^1 \rho_{12} P_s^1 .$$

All information on Σ_2 is contained in the reduced statistical operator ρ_2, which we obtain tracing out Σ_1:

$$\rho_2 = Tr_1(\rho'_{12}) = Tr_1\left(\sum_k P_k^1 \rho_{12} P_k^1\right) = \sum_k Tr_1\left(P_k^1 \rho_{12} P_k^1\right).$$

By using the properties of the Trace (1.1), it follows that

$$\rho_2 = \sum_k Tr_1\left(P_k^1 \rho_{12} P_k^1\right) = \sum_k Tr_1\left(P_k^1 \rho_{12}\right) = Tr_1\left(\sum_k P_k^1 \rho_{12}\right) = Tr_1 \rho_{12},$$

which is the exact reduced density operator one would have obtained if no measurement on Σ_1 had been performed.

Remark 4.5. It is not possible to distinguish whether a measurement on Σ_1 has been made by performing measurements only on Σ_2.

The following theorem [**40**] is closely related to 4.1: It is not possible to operate on a quantum system, extract some information from it, and then reconstruct the quantum system to its original state by using this information.

Theorem 4.3. (No classical coding)
Let $\mathcal{A}$ and $\mathcal{B}$ be *-algebras and let $C : \mathcal{B} \longrightarrow \mathcal{A}$ and $D : \mathcal{A} \longrightarrow \mathcal{B}$ be operations (coding and decoding), such that $C \circ D = id_{\mathcal{A}}$. Then,

$$\mathcal{B} \text{ is Abelian} \Rightarrow \mathcal{A} \text{ is Abelian}.$$

4.5. Quantum novelties

In the previous section, we have seen certain -somewhat strange- limitations that quantum operations are subject to. Let us now treat the other side of the coin: quantum mechanics allows for new surprising possibilities. In this section, we will not treat the really sensational novelties, such as very fast computation and secure cryptography. Here we shall treat quantum teleportation and quantum dense coding.

Example 4.5. Quantum dense coding
Quantum dense coding is a quantum communication protocol addressed to transmit 2 bits of information in a single qubit. The idea is the use of an entangled Bell state, e.g.,

$$|\psi^+\rangle = \frac{|0\rangle|1\rangle + |1\rangle|0\rangle}{\sqrt{2}}.$$

Alice keeps one particle, whereas the other is sent to Bob, who can perform one of the following actions:

1. an identity operation;
2. a state flip; namely, $|0\rangle \longleftrightarrow |1\rangle$;

3. a state-dependent phase shift; namely, a phase shift differing by π for the two qubits;
4. the two previous steps together.

This allows for the generation of all Bell states; namely,

$$|\phi_{+,-}\rangle = \frac{|0\rangle|0\rangle \pm |1\rangle|1\rangle}{\sqrt{2}}, \quad |\psi_{+,-}\rangle = \frac{|0\rangle|1\rangle \pm |1\rangle|0\rangle}{\sqrt{2}}.$$

Thus, if Bob sends back his particle to Alice, she can obtain two bits, by measuring the Bell state of the pair [**2, 40, 48**].

Example 4.6. Quantum teleportation
Suppose that Alice wishes to send to Bob the quantum state ρ over a (classical) telephone line. Theorem 4.3 states that, without further tools, this is impossible.

If Alice and Bob share an entangled pair, then Alice is indeed able to transfer her qubit to Bob [**2, 40, 48, 49**]. The destruction of the original state ρ cannot be avoided; otherwise Alice and Bob would have copied the state ρ, contradicting Theorem 4.1. This is the main reason for the name *teleportation*.

Chapter 5
Field Theoretical Methods

In this chapter, we consider the main lines that involve Field Theoretical Methods, as a bridge from theory to practical implementations (which physics are involved in the control of a quantum procedure and how they can be applied to perform it) [**3, 21, 41, 69**].

The exposition one can find in the literature of this subject is quite standard and slightly varies from different authors. We have chosen to base it on [**3**] for its simplicity and straightforwardness.

This chapter aims to set up the basis for comprehending the experiments described in Part 4, as well as to give a more applied justification of the axioms presented in the previous chapter (For example, compare Axiom 4 and Equation (5.11)).

5.1. Quantum Harmonic Oscillator

In quantum optics, a quantum harmonic oscillator provides a good model for a single mode of radiation confined in an optical cavity −the classical variables q and p corresponding to position and momentum, respectively, can be seen as the amplitudes of the magnetic and electric fields, respectively, in the quantum case−.

The Hamiltonian of a classical harmonic oscillator, described in terms of canonical position and momentum variables q and p, is

$$H(q,p) = \frac{1}{2}(\frac{p^2}{m} + m\omega^2 q^2),$$

where ω is its angular frequency and m is its mass.

We can safely make the assumption that $m = 1$ by choosing the appropriate units of measure.

Thus, the Hamiltonian becomes

$$H(q,p) = \frac{1}{2}(p^2 + \omega^2 q^2). \tag{5.1}$$

The solution of the Hamilton's equations of motion, which are

$$\frac{dq}{dt} = \frac{\partial}{\partial p}H(q,p) = p, \quad \frac{dp}{dt} = -\frac{\partial}{\partial q}H(q,p) = -\omega^2 q,$$

can be written in a more convenient form using a complex amplitude $z = \omega q + ip$:

$$\frac{dz}{dt} = -i\omega z, \quad z(t) = z(0)e^{-i\omega t}.$$

The Heisenberg approach to quantization consists in replacing the classical variables q, p by non-commuting self-adjoint operators, $\hat{q}$ and $\hat{p}$, acting on a Hilbert space $\mathcal{H}$.

Remark 5.1. $\hat{q} = q$ corresponds to the *position operator* and $\hat{p} = -i\hbar\frac{\partial}{\partial q}$ corresponds to the *momentum operator*. Also, note that the first term in (5.1) is the *kinetic energy* of the particle and the second term is the *potential energy*.

We want to find the energy levels E and its corresponding energy eigenstates $|\psi\rangle$, such that

$$\frac{1}{2}(\hat{p}^2 + \omega^2\hat{q}^2)|\psi\rangle = E|\psi\rangle.$$

The differential equation can be solved by standard means, leading to a family of solutions

$$|\psi_n\rangle = \psi_n(q) = \sqrt{\frac{1}{2^n n!}} \cdot \sqrt[4]{\frac{\omega}{\pi\hbar}} \cdot e^{-\frac{\omega q^2}{2\hbar}} H_n\left(\sqrt{\frac{\omega}{\hbar}}q\right), \quad n \in \mathbb{N} \cup \{0\}, \tag{5.2}$$

where $H_n(x)$ is the *Hermite polynomial* $H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}\left(e^{-x^2}\right)$. The corresponding energy levels are $E_n = \hbar\omega(n + \frac{1}{2})$.

Although straightforward, this spectral method solution is rather tedious. We will use the *ladder operator* method, which is due to Paul Dirac, and enables us to extract the energy eigenvalues without directly solving the differential equation. Moreover, it can be easily generalized to more complicated problems in Quantum Field Theory.

We want $\hat{p}$ and $\hat{q}$ to satisfy the following canonical commutation relation:

Definition 5.1. A canonical commutation relation (CCR) is of the form

$$[\hat{q}, \hat{p}] := \hat{q}\hat{p} - \hat{p}\hat{q} = i\hbar.$$

A normalized quantum complex amplitude is

$$\hat{a} = \frac{1}{\sqrt{2\hbar\omega}}(\omega\hat{q} + i\hat{p}).$$

Its hermitian conjugate $\hat{a}^\dagger$ satisfies the CCR $[\hat{a}, \hat{a}^\dagger] = 1$.

We also define the operator $\hat{n} = \hat{a}^\dagger\hat{a}$, and consider its set of normalized eigenvectors $|n\rangle$ satisfying [**3**]:

$$\hat{n}|n\rangle = n|n\rangle, n = 0, 1, 2, \ldots, \quad \langle n|n'\rangle = \delta_{n,n'}.$$

The CCR implies the following formulas:

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle, \quad \hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle, \quad n = 0, 1, 2, \ldots, \tag{5.3}$$

which justify the names annihilation, creation and particle number for the respective operators $\hat{a}, \hat{a}^\dagger, \hat{n}$.

Thus, one has the quantum Hamiltonian

$$\hat{H} = \hbar\omega\left(\hat{a}^{\dagger}\hat{a} + \frac{1}{2}\right), \tag{5.4}$$

which corresponds to the quantization of (5.1) by a simple calculation. However, the constant $\hbar\omega/2$ is irrelevant and we normally use as Hamiltonian for the harmonic oscillator the operator

$$\hat{H}_0 = \hbar\omega\hat{a}^{\dagger}\hat{a}. \tag{5.5}$$

Its eigenstates are $|n\rangle$ and its eigenvalues correspond to $E_n = \hbar\omega n$. The ground state $|0\rangle$ is called the vacuum.

Any classical observable $F(q,p)$ can be thought as $F(q,p) \equiv F(\alpha,\alpha^*) = \sum\limits_{k,l} c_{kl}(\alpha^*)^k\alpha^l$, where $\alpha = \frac{1}{\sqrt{2\hbar\omega}}(\omega q + ip)$.

In order to define its quantum counterpart

$$\hat{F} = \sum_{k,l} c_{kl}(\hat{a}^{\dagger})^k\hat{a}^l,$$

one has to consider the normal ordering that Wick's theorem [**21, 70**] guarantees, and which states that a string of creation and annihilation operators can be expressed as a sum of normal ordered terms using contractions.

The time evolution of $\hat{a}$ in Heisenberg picture is given by

$$\hat{a}(t) = e^{\left(\frac{i}{\hbar}\hat{H}_0 t\right)}\hat{a}e^{\left(-\frac{i}{\hbar}\hat{H}_0 t\right)}.$$

Differentiating both sides of the previous equation and using the CCR $[\hat{H}_0, \hat{a}] = -\hbar\omega\hat{a}$ one obtains [**3**]

$$\hat{a}(t) = e^{-i\omega t}\hat{a}.$$

Therefore, all quantum observables written in terms of the normal ordering evolve in Heisenberg picture in a similar way to their classic counterparts:

$$\hat{F}(t) = \sum_{k,l} c_{kl}e^{i\omega t(k-l)}(\hat{a}^{\dagger})^k\hat{a}^l.$$

5.1.1. The Weyl Unitary Operators. The position and momentum operators $\hat{q}$, $\hat{p}$ define a quantum phase space (in this case, a plane). There is a natural notion of translation operators, which are called displacement operators or Weyl unitaries.

Definition 5.2. A Weyl unitary operator is

$$\hat{W}(\alpha) = e^{\alpha\hat{a}^{\dagger} - \alpha^*\hat{a}}, \quad \alpha \in \mathbb{C}\ .$$

Remark 5.2. The notion of displacement comes from the fact that, using CCR, one can easily prove [**3**]

$$\hat{W}(\alpha)\hat{a}\hat{W}(\alpha)^{\dagger} = \hat{a} - \alpha,$$

which means that a *shift* in position and momentum occurs such that $\alpha = (\omega q + ip)/\sqrt{2\hbar\omega}$.

Properties 5.1. The Weyl unitaries have the following properties:

- $\hat{W}(0) = \mathbb{I}$
- $\hat{W}(-\alpha) = \hat{W}(\alpha)^\dagger$
- $\hat{W}(\alpha)\hat{W}(\beta) = e^{i\Im(\alpha\beta^*)}\hat{W}(\alpha+\beta)$, where $\Im(a)$ is the imaginary part of $a \in \mathbb{C}$.

In order to prove the third property one can use the operator identity

$$e^{\hat{A}}e^{\hat{B}} = e^{\left(\frac{1}{2}[\hat{A},\hat{B}]\right)}e^{\hat{A}+\hat{B}},$$

which is valid if the following condition holds:

$$[\hat{A},[\hat{A},\hat{B}]] = [\hat{B},[\hat{A},\hat{B}]] = 0.$$

Finally, one has the formal composition formulas

$$\hat{W}(\alpha) = e^{-\frac{|\alpha|^2}{2}}e^{\alpha\hat{a}^\dagger}e^{-\alpha^*\hat{a}} = e^{\frac{|\alpha|^2}{2}}e^{-\alpha^*\hat{a}}e^{\alpha\hat{a}^\dagger}. \tag{5.6}$$

5.1.2. Coherent States. Now, shifting the vacuum vector $|0\rangle$ by Weyl unitaries, we can obtain the following:

Definition 5.3. A family of coherent states (or exponential vectors) is defined as

$$|\alpha\rangle = \hat{W}(\alpha)|0\rangle, \quad \alpha \in \mathbb{C}\ .$$

Properties 5.2. Coherent vectors possess the following properties, which makes them interesting:

- They are linearly independent, yet not orthogonal eigenvectors of the operator $\hat{a}$, which is non-self-adjoint:

  $$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle, \quad \langle\alpha|\beta\rangle = e^{-\frac{|\alpha|^2}{2}-\frac{|\beta|^2}{2}+\alpha^*\beta}, \tag{5.7}$$

  forming an over-complete set, i.e.,

  $$\mathbb{I} = \frac{1}{\pi}\int_{\mathbb{C}} d^2\alpha|\alpha\rangle\langle\alpha|.$$

- The representation of $|\alpha\rangle$ in terms of particle number eigenvectors is

  $$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle.$$

  Note this implies that the probability distribution of the particle number is Poissonian with parameter $|\alpha|^2$:

  $$|\langle n|\alpha\rangle|^2 = e^{-|\alpha|^2}\frac{|\alpha|^{2n}}{n!} = p(n,|\alpha|^2).$$

- The mean value of a normally ordered operator $\hat{F}$ coincides with its classical counterpart:

  $$\langle\alpha|\hat{F}|\alpha\rangle = F(\alpha^*,\alpha) = \sum_{k,l}(\alpha^*)^k\alpha^l.$$

  Furthermore, a coherent state evolves into a coherent state, following the classical trajectory

  $$e^{\left(-\frac{i}{\hbar}\hat{H}_0 t\right)}|\alpha\rangle = |\alpha(t)\rangle, \quad \alpha(t) = e^{-i\omega t}\alpha.$$

- Considering the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$ and the position and momentum operators given by
$$(\hat{q}\psi)(x) = x\psi(x) \quad (\hat{p}\psi)(x) = -i\hbar\frac{d}{dx}\psi(x),$$
one can write the condition (5.7) $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$ and immediately obtain the linear differential equation (for $|\alpha\rangle = \psi(x) \in \mathcal{H}$):
$$\frac{d}{dx}\psi(x) = \frac{1}{\hbar}(ip - \omega(x-q))\psi(x),$$
with normalized solution (which has a similar form to (5.2), as expected)
$$\phi_\alpha(x) = (\pi\hbar)^{-1/4}e^{\left(\frac{i}{\hbar}px\right)}e^{\left(-\frac{\omega}{2\hbar}(x-q)^2\right)},$$
which is the wave function for the coherent state $|\alpha\rangle$. This is called the *Schrödinger position representation.*
One can obtain the momentum representation of the coherent state via the Fourier transform of $\phi_\alpha(x)$:
$$\tilde{\phi_\alpha}(v) = (\pi\hbar)^{-1/4}e^{\left(\frac{i}{\hbar}q(p-v)\right)}e^{\left(-\frac{(v-p)^2}{2\omega\hbar}\right)}.$$

- If we compute the corresponding Gaussian probability distributions $|\phi_\alpha(x)|^2$ and $\left|\tilde{\phi_\alpha}(v)\right|^2$, then for coherent states the Heisenberg uncertainty relation $\Delta\hat{q}\Delta\hat{p} \geq \hbar/2$ reaches the equality when there is symmetry between position and momentum, i.e., $\Delta\hat{q} = \sqrt{\hbar/2\omega}, \Delta\hat{p} = \sqrt{\hbar\omega/2}$. This means that the coherent vector gives the best quantum analogue of the classical state which is located at the point (q,p) in the phase-space.

As an example of the operativity involved, we shall indicate the proofs of the first two points:
The parametrization $\alpha = Re^{i\Theta}$, $d^2\alpha = R\,dR\,d\Theta$ and then integrating first over Θ leads to the desired result.
It is easy to obtain
$$(\hat{a}^\dagger)|0\rangle = \sqrt{n!}|n\rangle \tag{5.8}$$
and then
$$|\alpha\rangle = \hat{W}(\alpha)|0\rangle = e^{-\frac{|\alpha|^2}{2}}e^{(\alpha\hat{a}^\dagger)}|0\rangle,$$
where we used the formal composition formulas (5.6) applied to the state $|0\rangle$ using (5.3). Thus, one finally arrives at
$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle.$$

5.2. Quantum Bosonic Fields

The theory of quantum bosonic fields has two interpretations: The first treats it as a quantization of the classical macroscopic field theory (e.g. classical electrodynamics or acoustic waves in solids); the second, as a theory of many-body systems consisting of quantum particles, each of which is described by a suitable wave equation −this is called second quantization−. In this case Maxwell equations are treated as the Schrödinger equation for a single photon.

5.2.1. Quantization of Classical Fields. To begin with the theory of quantum bosonic fields from the first point of view, we will introduce the formalism of classical field theory, which is convenient for describing macroscopic properties and fundamental interactions (gravitational, electromagnetic, etc.) and for providing simplified continuous models of many-body systems (acoustic waves).

A classical field is a (generally multi-component) function $\phi(\boldsymbol{x};t)$ which satisfies some linear wave equation. For mathematical simplicity, we shall consider a real scalar field in a finite region $\Omega \subset \mathbb{R}^n, n = 1, 2, 3$ with specific boundary conditions.

The particular solutions (complex-valued) of the wave equation with periodic time dependence may be written as

$$u_k(\boldsymbol{x}, t) = e^{-i\omega(k)t} u_k(\boldsymbol{x}),$$

where k is a certain multi-index identifying the modes u_k. These modes satisfy the orthogonality conditions for some normalization constants c_k:

$$\int_\Omega d\boldsymbol{x}\ u_k^*(\boldsymbol{x}) u_l(\boldsymbol{x}) = c_k \delta_{kl}.$$

Remark 5.3. Applying the superposition principle, any solution f of the wave equation can be written in terms of modes and amplitudes $\alpha_k \in \mathbb{C}$

$$f(\boldsymbol{x}, t) = \sum_k (\alpha_k(t) u_k(\boldsymbol{x}) + \alpha_k^*(t) u_k^*(\boldsymbol{x})), \quad \alpha_k(t) = \alpha_k e^{-i\omega(k)t}.$$

Remark 5.4. As the wave equation is assumed to be linear, this suggests a quadratic dependence of the corresponding energy.

The energy of the field $f(\boldsymbol{x}, t)$ is given by the quadratic form

$$\mathcal{E}(f) = \int_\Omega d^r \boldsymbol{x} f(\boldsymbol{x}, t) \hat{D} f(\boldsymbol{x}, t), \quad \boldsymbol{x} \in \mathbb{R}^r,$$

where $\hat{D}$ is a differential operator whose form depends on the wave equation.

Remark 5.5. The modes u_k are eigenfunctions of $\hat{D}$:

$$\hat{D} u_k = \lambda_k u_k.$$

If we choose as normalization constants $c_k = \sqrt{\frac{\hbar\omega(k)}{2\lambda_k}}$ and use the superposition principle, then the energy can be written in terms of the complex amplitudes as

$$\mathcal{E}(f) = \sum_k \hbar\omega(k) \alpha_k^* \alpha_k.$$

The step to quantization now is to replace the complex amplitudes α_k, α_k^* by a set of independent annihilation and creation operators $\hat{a}_k, \hat{a}_k^\dagger$ satisfying a general form of CCR

$$[\hat{a}_k, \hat{a}_l^\dagger] = \delta_{kl}, \quad [\hat{a}_k, \hat{a}_l] = [\hat{a}_k^\dagger, \hat{a}_l^\dagger] = 0 \tag{5.9}$$

Thus, one defines the quantum Hamiltonian for the field in the bounded region as follows:

$$\hat{H}_F = \sum_k \hbar\omega(k) \hat{a}_k^\dagger \hat{a}_k$$

and the oscillations of the quantum amplitudes are

$$\hat{a}_k(t) = e^{\left(\frac{i}{\hbar}\hat{H}_F t\right)} \hat{a}_k e^{\left(-\frac{i}{\hbar}\hat{H}_F t\right)} = e^{-i\omega(k)t} \hat{a}_k,$$

and similarly for the case of a simple quantum harmonic oscillator with Hamiltonian $\hat{H}_0$.

Remark 5.6. This shows that a quantum field in a finite space region is equivalent to a possibly infinite family of independent quantum harmonic oscillators.

Now, the quantum analogue of the classical field is the operator-valued function

$$\hat{f}(\boldsymbol{x},t) = \sum_k (\hat{a}_k e^{-i\omega(k)t} u_k(\boldsymbol{x}) + \hat{a}_k^\dagger e^{i\omega(k)t} u_k^*(\boldsymbol{x})).$$

5.2.2. The Fock Space. For the quantized field, its Hilbert space can be treated formally as the Hilbert space of a family of harmonic oscillators; this is a tensor product of single-oscillator Hilbert spaces $\bigotimes_k \mathcal{H}_k$. When there is an infinite number of modes, the mathematical problem becomes subtle, so one constructs explicitly the appropriate Hilbert space, denoted by $\mathcal{F}$ and called Fock space.

Definition 5.4. The Fock Space $\mathcal{F}$ is spanned by a countable family of vectors which form an orthonormal basis denoted by $|\{n_k\}\rangle = |n_{k_1}, \dots, n_{k_m}\rangle$. That is, n_k is an arbitrary sequence of non-negative integer numbers with finitely many non-zero elements denoted by n_{k_j}.

Remark 5.7. The interpretation of this vector is a quantum state of the field, for which we observe n_k particles (excitations) corresponding to mode u_k. In this picture, called *particle number representation*, the modes u_k span the single-particle Hilbert space $\mathcal{H}_1$ and the particles are indistinguishable.

Formula (5.8) motivates

$$|n_{k_1}, \dots, n_{k_m}\rangle = \frac{(\hat{a}_{k_1}^\dagger)^{n_{k_1}}}{\sqrt{n_{k_1}!}} \cdot \frac{(\hat{a}_{k_2}^\dagger)^{n_{k_2}}}{\sqrt{n_{k_2}!}} \cdots \frac{(\hat{a}_{k_m}^\dagger)^{n_{k_m}}}{\sqrt{n_{k_m}!}} |0\rangle, \tag{5.10}$$

which can be used as a consistent definition of the creation operators $\hat{a}_k^\dagger$, which increase by one the number of particles in the state (mode) u_k.

Computing its adjoint one arrives at

$$\hat{a}_k|\dots, n_k, \dots\rangle = \sqrt{n_k}|\dots, n_k - 1, \dots\rangle.$$

Remark 5.8. It is easy to check that the CCR (5.9) are fulfilled and the states $|\{n_k\}\rangle$ are joint eigenvectors for the particle number operators $\hat{n}_k = \hat{a}_k^\dagger \hat{a}_k$:

$$\hat{n}_k|\{n_k\}\rangle = n_k|\{n_k\}\rangle.$$

Definition 5.5. We shall denote by $\mathcal{H}_N$ the subspace of $\mathcal{F}$ spanned by the vectors $|\{n_k\}\rangle$, where $N = \Sigma_k n_k$.

Then, the Fock space can be decomposed into a direct sum

$$\mathcal{F} = \bigoplus_{N=0}^{\infty} \mathcal{H}_N.$$

The subspace $\mathcal{H}_0$ is a one-dimensional ray generated by the vacuum $|0\rangle$.
The single-particle Hilbert space $\mathcal{H} \equiv \mathcal{H}_1$ is a complex space spanned by the vectors $\hat{a}_k^\dagger|0\rangle$.
This single-particle Hilbert space can be identified with a complex Hilbert space containing the linear combinations of the normalized modes $e_k(\boldsymbol{x})$:

$$\phi(\boldsymbol{x}) = \sum_k \phi_k e_k(\boldsymbol{x}), \quad e_k(\boldsymbol{x}) = \sqrt{\frac{2\lambda_k}{\hbar\omega(k)}} u_k(\boldsymbol{x}),$$

and its scalar product is given by

$$\langle\phi|\psi\rangle = \int_\Omega d\boldsymbol{x}\phi^*(\boldsymbol{x})\psi(\boldsymbol{x}) = \sum_k \phi_k^*\psi_k.$$

Finally, the N-particle Hilbert space $\mathcal{H}_N$ is spanned by the vectors $u_{k_1}u_{k_2}\cdots u_{k_N}$ −note that the same mode can appear many times and the order in which we perform the multiplication is irrelevant−. Therefore, the structure of $\mathcal{H}_N$ is the same as the N-fold symmetric tensor product. More precisely, if we introduce the permutation operator

$$S_\pi\phi_1(\boldsymbol{x}_1)\phi_2(\boldsymbol{x}_2)\cdots\phi_N(\boldsymbol{x}_N) = \phi_{\pi(1)}(\boldsymbol{x}_1)\phi_{\pi(2)}(\boldsymbol{x}_2)\cdots\phi_{\pi(N)}(\boldsymbol{x}_N), \quad \pi \in \mathfrak{S}_N,$$

then we have

$$\mathcal{H}_N = \frac{1}{N!}\sum_{\pi\in\mathfrak{S}_N} S_\pi(\mathcal{H}^{\otimes N}).$$

Remark 5.9. Thus, the Fock space can be viewed as an exponential of the underlying single-particle Hilbert space $\mathcal{H}$:

$$\mathcal{F}(\mathcal{H}) = \bigoplus_{N=0}^{\infty}\frac{1}{N!}\sum_{\pi\in\mathfrak{S}_N} S_\pi(\mathcal{H}^{\otimes N}).$$

5.2.3. Local Structure of Quantum Fields. By comparison of orthonormal bases of the Hilbert spaces which appear on both sides of the following equation can be seen that the following relation holds:

$$\mathcal{F}(\mathcal{H}\oplus\mathcal{K}) = \mathcal{F}(\mathcal{H})\otimes\mathcal{F}(\mathcal{K}).$$

Now, let's consider that the single-particle Hilbert space $\mathcal{H}$ possesses a natural local structure, due to the $\boldsymbol{x}$-dependence of its elements, which are complex wave functions $\phi(\boldsymbol{x})$, as stated in the previous section.

If we decompose the space region Ω into pairwise disjoint subsets

$$\Omega = \Omega_1 \sqcup \Omega_2 \sqcup \cdots \sqcup \Omega_k,$$

then any wave function can be written as an orthogonal sum of wave functions, each one localized in the corresponding Ω_j.

This generates a corresponding decomposition of the wave function:

$$\phi(\boldsymbol{x}) = \phi_1(\boldsymbol{x})\oplus\phi_2(\boldsymbol{x})\oplus\cdots\oplus\phi_k(\boldsymbol{x})$$

and, consequently, of the Hilbert space:

$$\mathcal{H} = \mathcal{K}_1\oplus\mathcal{K}_2\oplus\cdots\oplus\mathcal{K}_k.$$

Thus, we obtain the following decomposition of the Fock space:

$$\mathcal{F}(\mathcal{H}) = \mathcal{F}(\mathcal{K}_1)\otimes\mathcal{F}(\mathcal{K}_2)\otimes\cdots\otimes\mathcal{F}(\mathcal{K}_k). \tag{5.11}$$

Remark 5.10. The physical meaning of equation (5.11) is that the quantum field, localized in a certain subset $\Omega_j \subset \Omega$, is a physical subsystem of the total system, which is localized in Ω.

Example 5.1. *Photons, phonons, spin and statistics.*
The most important example of a quantized field is an electromagnetic field in vacuum.
We will begin by considering Maxwell's classical field equations for the electric and magnetic fields $\boldsymbol{E}(\boldsymbol{x},t), \boldsymbol{B}(\boldsymbol{x},t)$, which we shall write in a more compact notation using a complex vector field $\boldsymbol{Z}(\boldsymbol{x},t) = \boldsymbol{E}(\boldsymbol{x},t) + i\boldsymbol{B}(\boldsymbol{x},t)$.

Maxwell equations then are

$$\nabla \cdot \boldsymbol{Z} = 0, \quad \frac{\partial}{\partial t}\boldsymbol{Z} = -ic\nabla \times \boldsymbol{Z}.$$

Introducing periodic boundary conditions in a cubic box L^3, we can find solutions of Maxwell equations in terms of plane waves:

$$\boldsymbol{Z}(\boldsymbol{x},t) = \boldsymbol{Z}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega(\boldsymbol{k})t)},$$

with $\boldsymbol{k}$ and $\boldsymbol{Z}_0$ satisfying the orthogonality conditions

$$\boldsymbol{k} \cdot \boldsymbol{Z}_0 = 0, \quad ic\boldsymbol{k} \times \boldsymbol{Z}_0 = \omega(\boldsymbol{k})\boldsymbol{Z}_0.$$

The values that $\boldsymbol{k}$ can take are discrete: $\boldsymbol{k} = \frac{2\pi}{L}(m_1, m_2, m_3), m_l \in \mathbb{Z}$ and the dispertion law $\omega(\boldsymbol{k}) = c\,|\boldsymbol{k}|$ holds.
As the wave vector $\boldsymbol{k}$, the electric field and the magnetic field are mutually orthogonal, we have two possible polarizations of the plane wave, which we shall denote by the index $\lambda = \pm 1$.
Finally, the energy is given by the quadratic form

$$\mathcal{E} = \frac{1}{8\pi}\int d\boldsymbol{x}(\boldsymbol{E}^2 + \boldsymbol{B}^2) = \frac{1}{8\pi}\int d\boldsymbol{x}\boldsymbol{Z}^* \cdot \boldsymbol{Z},$$

so that we can apply the quantization procedure mentioned above and introduce annihilation and creation operators $\hat{a}_{\boldsymbol{k},\lambda}, \hat{a}^\dagger_{\boldsymbol{k},\lambda}$ for plane waves with given wave vectors and polarizations.
The quantized electric field we obtain is

$$\hat{\boldsymbol{E}}(\boldsymbol{x}) = i\sum_{\boldsymbol{k},\lambda}\left(\frac{2\pi\hbar c\,|\boldsymbol{k}|}{L^3}\right)^{1/2} \boldsymbol{e}_{\boldsymbol{k},\lambda}(e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\hat{a}_{\boldsymbol{k},\lambda} - e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\hat{a}^\dagger_{\boldsymbol{k},\lambda}).$$

Remark 5.11. In order to quantize the oscillations of interacting atoms, ions, or molecules in a solid state, we associate with any normal oscillation mode (described by the wave vector $\boldsymbol{k}$ and the index α corresponding to the possible kinds of oscillations e.g. three polarizations of acoustic modes, different branches of optical modes...) annihilation and creation operators $\hat{a}_{\boldsymbol{k},\alpha}$ and $\hat{a}^\dagger_{\boldsymbol{k},\alpha}$ and a dispersion law $\omega(\boldsymbol{k},\alpha)$. Note that in contrast to electromagnetic waves, the number of modes is finite and proportional to the volume. The particles corresponding to this picture are called phonons.

Remark 5.12. The multi-particle structure of the Fock space suggests its direct application to the description of many-body systems which consist of particles not associated with macroscopic classical fields in an obvious way.
However, from relativistic quantum field theory, it follows that only particles with an integer spin $S = 0, 1, 2, \ldots$ called bosons can be described with the mentioned formalism. Examples of such particles are photons, pions, W,Z-bosons.
For particles with a spin $S = 1/2, 3/2, \ldots$, called fermions (e.g. electrons, protons, neutrons), Pauli's exclusion principle must be satisfied, and this implies that the possible eigenvalues of the particle number operators $\hat{n}_k$ should be equal to 0 or 1.

In order to fit this requirement into our theory, a second quantization procedure must be made, this time involving canonical anti-commutation relations instead of CCR.

Remark 5.13. We say that bosons obey Bose-Einstein statistics, while fermions are subjected to Fermi-Dirac statistics.

5.3. 2nd Quantization of Fermions

Remark 5.14. Pauli exclusion principle.
In contrast to bosons, two or more fermions cannot macroscopically occupy a single quantum state. Therefore, a wave function of a fermion has no classical meaning of a measurable macroscopic field. However, we can describe many fermion systems in terms of the second quantization, which is very similar to the Bosonic one. The procedure is as follows:
Choosing again a certain collection of modes $\{u_k\}$, which are solutions of a single-fermion Schrödinger equation, we can define the corresponding annihilation and creation operators $\hat{c_k}, \hat{c_k}^\dagger$. To fulfill the requirement that the particle number operators must have only $0, 1$ eigenvalues that can be achieved imposing the following canonical anti-commutation relations (CAR):

$$\{\hat{c}_k, \hat{c}_l^\dagger\} = \delta_{kl}, \quad \{\hat{c}_k, \hat{c}_l\} = \{\hat{c}_k^\dagger, \hat{c}_l^\dagger\} = 0,$$

being $\{\hat{A}, \hat{B}\} = \hat{A}\hat{B} + \hat{B}\hat{A}$. Therefore, paricle number operators satisfy the relations

$$\hat{n}_k^2 = \hat{n}_k, \quad [\hat{n}_k, \hat{n}_l] = 0.$$

Example 5.2. Fock Space
The minimal Hilbert space, called Fermionic Fock space and denoted by $\mathcal{F}^a$, that supports the structure of operators described in this section is spanned by the Fock vectors that are the joint eigenvectors of all $\hat{n}_k$:

$$\hat{n}_k|\{n_k\}\rangle = n_k|\{n_k\}\rangle, \quad n_k \in \{0, 1\}.$$

Now, for any $N = 0, 1, 2, \ldots$ we define a subspace of $\mathcal{F}^a$, spanned by the vectors $|\{n_k\}\rangle$ with $\sum_k n_k = N$ and we shall denote it $\mathcal{H}_N^a$.

De decomposition of the Fock space into a direct sum reads

$$\mathcal{F}^a = \bigoplus_{N=0}^{\infty} \mathcal{H}_N^a.$$

However, $\mathcal{H}_N^a$ is the antisymmetric tensor product spanned by the antisymmetrization of the formal products $u_{k_1} u_{k_2} \cdots u_{k_N}$ and now each mode can only appear once.

The subspace $\mathcal{H}_0^a$ is a one-dimensional ray generated by the vacuum $|0\rangle$. The single-particle Hilbert space $\mathcal{H} = \mathcal{H}_1^a$ is spanned by the modes u_k.

Its local structure is described by the same relation as for the bosonic case:

$$\mathcal{F}(\mathcal{H}^a \oplus \mathcal{K}^a) = \mathcal{F}(\mathcal{H}^a) \otimes \mathcal{F}(\mathcal{K}^a).$$

Remark 5.15. Thus, we have seen that the structure is similar for Bosons and Fermions: The Fock space of a direct sum turns into the tensor product of Fock spaces.

However, note that Symmetric Algebra corresponds to Bosonic particles, whereas Exterior Algebra corresponds to Fermions.

Part 3

Quantum Information Processing

Chapter 6
Quantum Algorithms and Computing

In this chapter we introduce a circuit model to support the representation of a quantum computer and present some examples. We present algorithms which are subroutines used in many (more specific) algorithms, like the Quantum Fourier Transform, and we also give some complete algorithms, i.e., programs which end with a measurement process. We make the distinction between categoric algorithms (which give always the same result with certainty) and probabilistic algorithms (which may not always give the same result) and analyze it consequences.

In particular, we present the algorithms which have had the most serious implications (Deutsch-Josza, Simon, Shor, Grover, Phase estimation) and we treat the more general framework of the Hidden Subgroup Problem. We finally analyze the dependencies of the main quantum algorithms known. [**31, 32, 48, 44, 50, 51, 56, 65**].

6.1. Introduction

In the 1980s, Feynman was the first to consider quantum mechanics from a computational point of view [**24**]; he observed that the simulation of quantum mechanical systems on a classical computer seemed to require an increase in complexity which was exponential in the size of the system. To circumvent this exponential overhead, he asked if it was possible to design a universal *quantum* computer [**25**]. Deutsch [**18**] was the first to exhibit a concrete computational task which admitted a quantum algorithm strictly more efficient than the best classical algorithm solving the same problem, as we shall see in this chapter. The most striking demonstration of the computational power of quantum computers was given by Peter Shor [**60**] in 1994, who exhibited efficient quantum algorithms for factoring integers and for computing the discrete logarithm and has definitely marked a milestone in this field, since many cryptographic systems nowadays rely on the difficulty of these problems.

6.2. Circuit Model

One can describe a classical computer by means of a circuit which takes as input a string of bits from $\{0,1\}^n$, processes them by a succession of logical gates such as NOT, OR, AND, NAND... and produces output bits, which can be viewed as Boolean functions $f : \{0,1\}^n \longrightarrow \{0,1\}$.
According to Feynman's model, a quantum computer obeys quantum mechanics, rather than Maxwell physics. This has important implications in the context of computation:

- The *states* describing the machine are wave functions: Each basic unit of computation -qubit- can be thought as a two-dimensional complex vector of norm 1 in a Hilbert space with basis $\{|0\rangle, |1\rangle\}$. The basis states can be thought as the states of a classical bit, $0, 1$.
- The *dynamics* governing the evolution of the state in time is *unitary.* A unitary matrix transforms the state at a certain time to the state at a later time.
- A second dynamical ingredient is *measurement.* The observation of a system changes it. In the context of a quantum algorithm a measurement can be thought of as a projection onto the computational basis.

We will denote by $\boldsymbol{U}^{(n)}$ the set of unitary matrices of dimension 2^n ($UU^\dagger = \mathbb{I}_{2^n}$). With the standard multiplication of matrices, $\boldsymbol{U}^{(n)}$ is a group, and its elements can be viewed as q-computations of order n: Transformations of n qubits into n qubits.

The group structure of $\boldsymbol{U}^{(n)}$ determines the following properties of q-computations:

- *Identity.* $\mathbb{I}_{2^n} \in \boldsymbol{U}^{(n)}$. The identity matrix of dimension 2^n is a q-computation of order n.
- *Composition.* If $U, V \in \boldsymbol{U}^{(n)}$, then $V \cdot U \in \boldsymbol{U}^{(n)}$. The composition of two q-computations of order n is a q-computation of order n.
- *Reversibility.* If $U \in \boldsymbol{U}^{(n)}$, then $U^{-1} = U^\dagger \in \boldsymbol{U}^{(n)}$. The inverse of a q-computation of order n is a q-computation of order n.

Given this model, it is not even clear if such quantum computer is able to perform classical computations. For instance, some elementary gates (NAND, OR...) are not reversible (2 inputs, a single output). Nevertheless, the question of reversibility in classical computation has been studied and has been established that with a polynomial overhead in the number of gates and bits used, classical computation can be made reversible [**32**].

Remark 6.1. q-computations of order n are vastly more abundant than classical reversible computations of the same order. This is already clear for $n = 1$, where the only classical reversible computation is NOT, whereas q-computations of 1 qubit depend on continuous parameters:

It is easy to see that they must be of the form

$$U = e^{i\alpha} \begin{pmatrix} u_0 & u_1 \\ -u_1^* & u_0^* \end{pmatrix}, \quad \alpha \in \mathbb{R}, u_0, u_1 \in \mathbb{C}\ , \quad |u_0|^2 + |u_1|^2 = 1.$$

In fact, any $U \in \boldsymbol{U}^{(1)}$ has this form and $\det U$ lies on the unit circle. Hence, omitting the global phase factor, we can assume a q-computation to be represented by $\begin{pmatrix} u_0 & u_1 \\ -u_1^* & u_0^* \end{pmatrix}$.

Note that the constraint $|u_0|^2 + |u_1|^2 = 1$ implies that $\exists! \theta \in [0, \pi]$ such that $|u_0| = \cos\frac{\theta}{2}$ and $|u_1| = \sin\frac{\theta}{2}$. For further convenience, we shall represent them as $u_0 = e^{-i\lambda}\cos\frac{\theta}{2}, u_1 = -e^{i\mu}\sin\frac{\theta}{2}$, $\lambda, \mu \in \mathbb{R}$.

Example 6.1. In this example we consider several particular cases of q-computations involving 1 qubit:

(a) Pauli matrices.
They are self-adjoint, $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbb{I}_2$. The Pauli matrix σ_x corresponds to classical

NOT operator[1]:

$$\sigma_x|0\rangle = |1\rangle \quad \sigma_x|1\rangle = |0\rangle.$$

In terms of arithmetic modulo 2, we have $\sigma_x|j\rangle = |j+1\rangle$

(b) Hadamard matrix.

The matrix $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is self-adjoint and $H^2 = \mathbb{I}_2$

(c) Phase shift matrices.

They are matrices of the form $S_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{pmatrix}$.

As particular cases, we define $S = S_{\pi/2}$ and $T = S_{\pi/4}$. Note that $\sigma_x = S^2 = T^4$, sometimes written $S = \sqrt{NOT}$.

6.2.1. The CNOT gate. The most important quantum gate of order 2 is probably the controlled-not CNOT gate. It takes 2 qubits $|a\rangle, |b\rangle$ and outputs the first qubit $|a\rangle$ and flips the second if the first was in the state $|1\rangle$: $|a \oplus b\rangle$. More precisely, it acts on the basis vectors as

$$|a\rangle|b\rangle \xrightarrow{CNOT} |a\rangle|a \oplus b\rangle, \quad a, b \in \{0, 1\}.$$

Thus, its matrix representation is

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

It turns out that the set of three gates $\{T, H, CNOT\}$ is universal, in the sense that any unitary transformation can be approximated to arbitrary precision by a sequence of gates from this set. This is an analogous result from the one in classical computation, where any classical algorithm can be implemented by means of the NAND and FANOUT gates, which shall be described in Remark 6.2.

6.2.2. The Toffoli gate. This gate transforms 3 qubits and is given by its matrix representation or, alternatively, its action onto the basis vectors:

$$\mathcal{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad |a\rangle|b\rangle|c\rangle \xrightarrow{\mathcal{T}} |a\rangle|b\rangle|c \oplus ab\rangle.$$

This gate is also unitary and it flips the last qubit if the first two are in the state $|11\rangle$.

Remark 6.2. A circuit consisting of Toffoli gates can simulate any classical circuit: Indeed, since any Boolean function can be implemented by a circuit of NAND and FANOUT gates, it is sufficient to show how to implement those.

[1] If we identify the classical states 0 and 1 with the computational basis $\{|0\rangle, |1\rangle\}$, the Pauli matrix σ_x performs the classical NOT operation, since it applies the corresponding permutation. Note that, in general, a quantum computation does not have a classical analogue.

The FANOUT gate copies 1 bit into 2 bits. This can be done by means of

$$\mathcal{T}(|1\rangle|a\rangle|0\rangle) = |1\rangle|a\rangle|a\rangle.$$

The NAND gate outputs 0 if, and only if, both input bits are 1. In other words, it acts as $NAND(a,b) = 1 \oplus ab$. Thus, it can be implemented as

$$\mathcal{T}(|a\rangle|b\rangle|1\rangle) = |a\rangle|b\rangle|1 \oplus ab\rangle.$$

Since classical reversible computation is just a permutation on the bit strings of its input, it is in particular unitary.

The extra qubits we need to add to make the computation reversible are called *ancillas*.

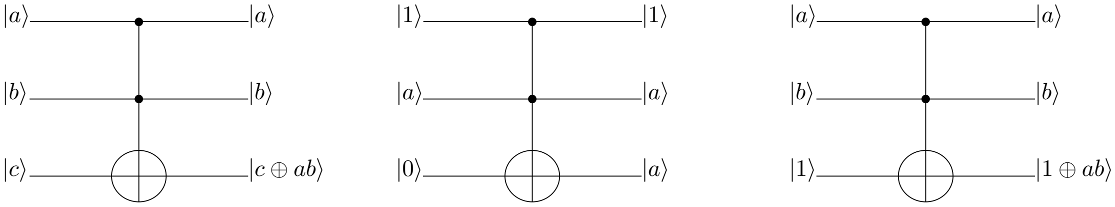

FIG. 6.1. Circuit representation of the Toffoli gate and its implementation of the FANOUT and NAND gates.

Remark 6.3. As a result, quantum computation is at least as strong as classical computation.

6.3. Quantum Fourier Transform

In this section we will briefly review some facts on the classical Fourier Transform on Abelian groups and we will define the quantum Fourier Transform, which is a key ingredient in many quantum algorithms.

Let $(G,+)$ be a finite Abelian group and $(\mathcal{U},\times)$ the multiplicative group of complex numbers with modulus 1.

Definition 6.1. A character on G is a function $\chi : G \longrightarrow \mathcal{U}$ such that

$$\forall g,h \in G \quad \chi(g+h) = \chi(g)\chi(h).$$

Remark 6.4. In the Abelian case, the set of all characters on G forms a group, $\tilde{G}$, which is isomorphic to G. We fix an isomorphism $\phi : G \longrightarrow \tilde{G}$ and denote by $\chi_g := \phi(g)$.

Lemma 6.1. (Schur's orthogonality lemma)
For every $h, h' \in G$,

$$\frac{1}{|G|}\sum_{g\in G} \chi_h(g)\chi_{h'}(g)^* = \delta_{hh'}.$$

Definition 6.2. (Abelian Fourier Transform)
Let $f : G \longrightarrow X$ be a function defined on G and taking its values in some $\mathbb{C}$ -vector space X

(Typical choices of X are $G, \mathbb{Z}, \mathbb{R}$, with the structure of $\mathbb{C}$ -vector space). The Fourier coefficients of f are

$$\hat{f}(x) = \frac{1}{\sqrt{|G|}} \sum_{y \in G} \chi_x(y) f(y).$$

The Fourier transform of f is the function

$$\hat{f} : x \mapsto \frac{1}{\sqrt{|G|}} \sum_{y \in G} f(y) \chi_x(y).$$

Remark 6.5. Let $M = |G|$ and $G = \{e_1, \dots e_M\}$. Then the classical Fourier Transform can be viewed as a linear transform and be represented by a matrix multiplication: We multiply the vector $\boldsymbol{f} = (f(e_1), \dots, f(e_M))$ by the Fourier Transform matrix

$$F = \frac{1}{\sqrt{|G|}} \begin{pmatrix} \chi_1(e_1) & \chi_1(e_2) & \cdots & \chi_1(e_M) \\ \chi_2(e_1) & \chi_2(e_2) & \cdots & \chi_2(e_M) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_M(e_1) & \chi_M(e_2) & \cdots & \chi_M(e_M) \end{pmatrix},$$

where we denoted $\chi_i = \chi(e_i)$ the characters of $\tilde{G}$.

From Schur's orthogonality lemma, it follows that F is unitary ($FF^\dagger = \mathbb{I}_M$). Thus, we obtain the vector of Fourier coefficients $\hat{\boldsymbol{f}} = F\boldsymbol{f}$. It is well known that this computation can be performed in $O(M \log M)$ elementary operations, using the Fast Fourier Transform (FFT) whereas naive matrix multiplication takes $O(M^2)$ elementary operations.

Definition 6.3. The Quantum Fourier Transform over an Abelian group G of cardinality M is the unitary operation:

$$QFT : |x\rangle \mapsto \frac{1}{\sqrt{M}} \sum_{y \in G} \chi_y(x) |y\rangle.$$

With any function $f : G \longrightarrow X$ we associate the state

$$|f\rangle = \frac{1}{\sqrt{M}} \sum_{x \in G} |x\rangle |f(x)\rangle,$$

and the state associated with the Fourier transform $\hat{f}$ of f is

$$|\hat{f}\rangle = \frac{1}{M} \sum_{x,y \in G} \chi_y(x) |y\rangle |f(x)\rangle.$$

We say that $\hat{f}$ is obtained by quantum Fourier sampling (QFS).

Remark 6.6. Note that quantum Fourier sampling can be thought of as the algorithm of Fig. 6.2. A measurement of the state $|\hat{f}\rangle$ will output $y \in G$ and $z \in X$ with probability proportional to $\left| \sum_{\substack{x \in G \\ z = f(x)}} \chi_y(x) \right|^2$. These amplitudes can be estimated later by repeated sampling.
$|\hat{f}\rangle$ gives us some *global* information on the function f through its Fourier coefficients and we will be able to make the best use of it to distinguish functions with very different Fourier coefficients.

Example 6.2. To end this section, we show how the state $|\hat{f}\rangle$ can be constructed. Sampling from this state is the basis of several algorithms we shall show in the next section.

We first need to consider the Black-Box Model: Suppose we are given a function $f : \{0,1\}^n \longrightarrow \{0,1\}^m$. f can be given to us in several formats (a table of values, an algorithm, mathematical characterization...). To abstract away any special feature f might have, we place ourselves in the black-box model of computation, in which f is given as a special gate U_f that we can use in our circuit. However, it is, in general, not reversible (e.g. f being a constant function). To make it reversible, we introduce another input $y \in \{0,1\}^m$. The box U_f acts as follows:

$$|x\rangle|y\rangle \xrightarrow{U_f} |x\rangle|y \oplus f(x)\rangle$$

and its a permutation of the $(n+m)$-bit strings; thus, reversible.

Now, let us assume that $|G| = M = 2^n$, $|X| = 2^p$ are powers of 2 and that we input the state $|0^{\otimes n}\rangle|0^{\otimes p}\rangle$ to the circuit of Fig. 6.2. We already know that $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. After the Hadamard transform on n qubits, the state is

$$(H^{\otimes n}|0^{\otimes n}\rangle)|0^{\otimes p}\rangle = \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right)^{\otimes n} |0^{\otimes p}\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle|0^{\otimes p}\rangle,$$

So we have prepared a quantum superposition of all possibilities of the first entry and $|0^{\otimes p}\rangle$ in the second. After applying the black box U_f we obtain the state $\sum\limits_{x\in\{0,1\}^n} |x\rangle|f(x)\rangle$, which, performing the QFT on the first n qubits gives us the final state

$$|\hat{f}\rangle = \frac{1}{M} \sum_{x,y\in G} \chi_y(x)|y\rangle|f(x)\rangle.$$

We shall denote this transformation D_f. Note that D_f uses only one gate U_f, yet it computes the QFT over all input possibilities (a quantum superposition of these).

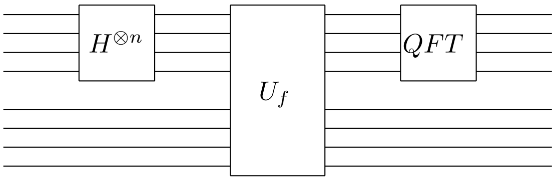

FIG. 6.2. The transformation D_f (Quantum Fourier sampling) in the general case.

Remark 6.7. In the following algorithms, we shall implement U_f by means of actual gates (CNOT, Hadamard...).

Example 6.3. The QFT over $\mathbb{Z}_2^n$.
Let $G = (\mathbb{Z}_n^2, \oplus)$. We are going to show that the QFT over G is just the tensor product of Hadamard gates, $H^{\otimes n}$.

Indeed, the action of H on a n-qubit basis state $|x\rangle = |x_1 \dots x_n\rangle$ on qubit i is

$$H|x_i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_i}|1\rangle) = \frac{1}{\sqrt{2}} \sum_{y_i\in\{0,1\}} (-1)^{x_i \cdot y_i}|y_i\rangle.$$

Applying this to $H^{\otimes n}$ we obtain

$$H^{\otimes n}|x\rangle = \frac{1}{\sqrt{2^n}} \sum_{y\in\{0,1\}^n} (-1)^{x\cdot y}|y\rangle,$$

where $x \cdot y = \sum_i x_i y_i \mod 2$.

Since the characters of $\mathbb{Z}_2^n$ are simply the applications $\chi_y(x) = (-1)^{x\cdot y}, y \in \mathbb{Z}_2^n$, it follows that $H^{\otimes n}$ implements the QFT over $\mathbb{Z}_2^n$.

Example 6.4. The QFT over $\mathbb{Z}_M$.
Another important group is the cyclic group of M elements $\mathbb{Z}_M$. We shall denote by $\omega = e^{i2\pi/M}$ an Mth primitive root of unity. The group of characters of $\mathbb{Z}_M$ is constituted of the M maps

$$\chi_i : \begin{array}{ccc} \mathbb{Z}_M & \longrightarrow & \mathcal{U} \\ k & \longmapsto & \omega^{i\cdot k} \end{array}, \quad 0 \leq i \leq M-1.$$

Hence, the QFT over $\mathbb{Z}_M$ is the unitary operation

$$QFT : |x\rangle \mapsto \frac{1}{\sqrt{M}} \sum_{y\in\mathbb{Z}_M} \omega^{x\cdot y}|y\rangle.$$

For simplicity, we shall assume $M = 2^n$. We shall introduce the notation $y = y_0 + 2y_1 + \ldots + 2^{n-1}y_{n-1}$ and $.x_i x_{i-1}\ldots x_0 = x_i/2 + x_{i-1}/4 + \ldots + x_0/2^{i+1}$. So, one computes

$$\sum_{y\in\mathbb{Z}_M} \omega^{x\cdot y}|y\rangle = \sum_{\substack{y_i\in\{0,1\}\\ i\in\{0\ldots n-1\}}} \omega^{xy_0}\omega^{2xy_1}\cdots\omega^{2^{n-1}xy_{n-1}}|y_0,\ldots,y_{n-1}\rangle$$

$$= \bigotimes_{i=0}^{n-1}\left(\sum_{y_i\in\{0,1\}} \omega^{2^i xy_i}|y_i\rangle\right) = \bigotimes_{i=0}^{n-1}\left(|0\rangle + \omega^{2^i x}|1\rangle\right).$$

Note that we can write $\omega^{2^i x} = e^{i2\pi 2^{i-n}x} = e^{i2\pi .x_{n-1-i}\ldots x_0}$.

This means that with an appropriate phase shift on each qubit we can successfully implement the QFT on $\mathbb{Z}_M$: More precisely, with the use of the phase shifts

$$R_d = \begin{pmatrix} 1 & 0 \\ 0 & e^{-i\frac{2\pi}{2^d}} \end{pmatrix},$$

the QFT is implemented as shown in Fig. 6.3 using $\frac{1}{2}n(n+1) = O(\log^2 M)$ gates.

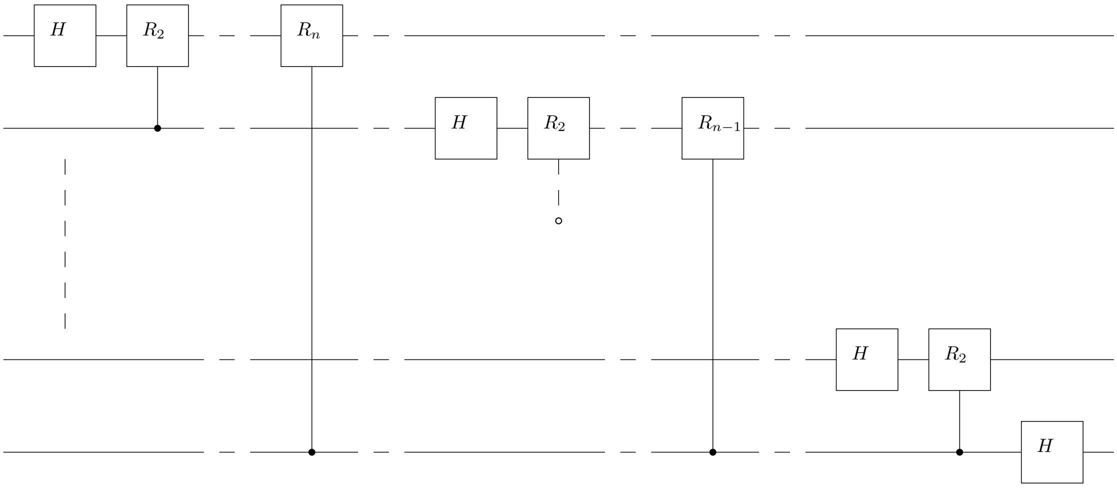

FIG. 6.3. The QFT over $\mathbb{Z}_m$.

6.4. Some Quantum Algorithms

In this section, we begin by showing two algorithms which make an elaborate use of the QFT over the group $\mathbb{Z}_2^n$: Deutsch-Josza and Simon's algorithms. Then we proceed to introduce Grover's search algorithm for an unstructured database. Later on, we move to Shor's algorithm for period finding and illustrate how the problem of factorization can be reduced to period finding in polynomial time. Finally we describe Kitaev's algorithm for phase estimation. Those two algorithms require the use of the QFT over the group $\mathbb{Z}_M$ or its inverse.

6.4.1. The Deutsch-Josza Algorithm. Consider the following problem [**18**] :
Input: An integer n and a Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$.
Assumption: f is either constant or balanced, i.e.,

$$\#\{x : f(x) = 0\} = \#\{x : f(x) = 1\} = 2^{n-1}.$$

Output: Constant or balanced.

A classical deterministic algorithm will need at least 2^{n-1} queries to f to solve the problem since we can always find a function which is constant on any 2^{n-1} x's in $\{0,1\}^n$ and we can choose its value on the remaining x's to make it constant or balanced.

However, this problem can be solved using exactly *one* query to U_f by means of the circuit D_f introduced in Fig. 6.2 with a slight modification. We will use the notation $\overline{f(x)} = f(x) \oplus 1$ for the complement of $f(x)$.

When performing the QFT we obtain:

- If f is constant, then

$$|\widehat{f}\rangle = \frac{1}{2^n} \sum_{x,y\in\mathbb{Z}_2^n} (-1)^{x\cdot y} |y\rangle |f(x)\rangle$$

$$= \frac{1}{2^n} \sum_{y\in\mathbb{Z}_2^n} \left(\sum_{x\in\mathbb{Z}_2^n} (-1)^{x\cdot y} \right) |y\rangle |f(0)\rangle = |0^{\otimes n}\rangle |f(0)\rangle,$$

  since $\sum_{x\in\mathbb{Z}_2^n} (-1)^{x\cdot y} = 2^n \delta_{y,0}$. An analogous calculation gives

$$|\widehat{\overline{f}}\rangle = |0^{\otimes n}\rangle |\overline{f(0)}\rangle.$$

- On the other hand, if f is balanced, then

$$|\widehat{f}\rangle = \frac{1}{2^n} \sum_{x,y\in\mathbb{Z}_2^n} (-1)^{x\cdot y} |y\rangle |f(x)\rangle$$

$$= \frac{1}{2^n} \sum_{y\in\mathbb{Z}_2^n} \left(|0^{\otimes n}\rangle \sum_{x\in\mathbb{Z}_2^n} |f(x)\rangle + \sum_{x,y\in\mathbb{Z}_2^n\setminus\{0\}} (-1)^{x\cdot y} |y\rangle |f(x)\rangle \right).$$

  Similarly,

$$|\widehat{\overline{f}}\rangle = \frac{1}{2^n} \sum_{y\in\mathbb{Z}_2^n} \left(|0^{\otimes n}\rangle \sum_{x\in\mathbb{Z}_2^n} |\overline{f(x)}\rangle + \sum_{x,y\in\mathbb{Z}_2^n\setminus\{0\}} (-1)^{x\cdot y} |y\rangle |\overline{f(x)}\rangle \right).$$

Observe that, for a balanced f, $\sum_x |f(x)\rangle = \sum_x |\overline{f(x)}\rangle$, so $|\widehat{f}\rangle - |\widehat{\overline{f}}\rangle$ has zero amplitude on all states which have their first register in the basis state $|0^{\otimes n}\rangle$.
On the other hand, for a constant f, $|\widehat{f}\rangle - |\widehat{\overline{f}}\rangle = |0^{\otimes n}\rangle(|f(x)\rangle - |\overline{f(x)}\rangle)$.

It is now clear that it is sufficient to measure the first register of $|\widehat{f}\rangle - |\widehat{\overline{f}}\rangle$. Then, if the outcome is $|0^{\otimes n}\rangle$ it means that f is constant; otherwise it is balanced.

We present the final circuit for solving the problem in Fig. 6.4

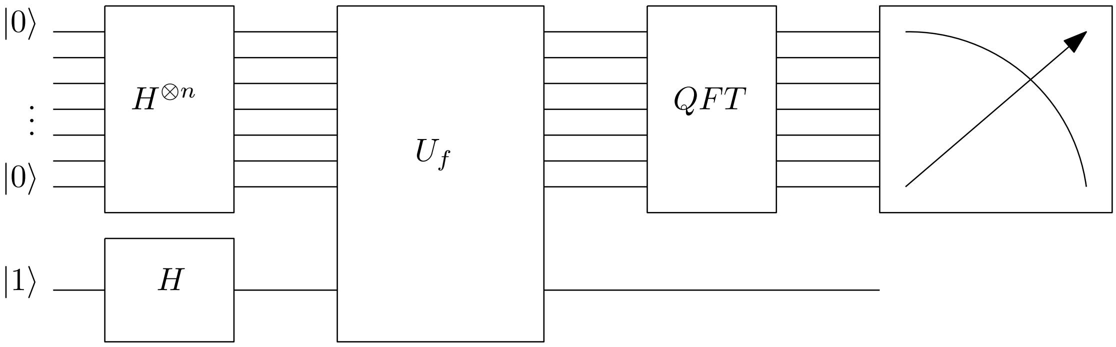

FIG. 6.4. The Deutsch-Josza circuit. The final box represents a measurement.

Remark 6.8. This algorithm is categoric -or exact- in the sense that it always produces the same result. The following algorithms' output is probabilistic; i.e., it might not always give the same result.

6.4.2. Simon's Algorithm. This is the first quantum algorithm that presents exponential advantage over the best classical probabilistic algorithm. It studies the periodicity of functions defined on $\mathbb{Z}_2^n$. We consider the problem:

Input: A function $f : \mathbb{Z}_2^n \longrightarrow \{0,1\}^n$.
Assumption: f is periodic: $\exists a \in \mathbb{Z}_2^n : \ \forall x, y \in \mathbb{Z}_2^n, \qquad y = x \oplus a \Rightarrow f(x) = f(y)$.
Output: a.

Since f does not have any particular structure apart from being periodic, the best classical probabilistic algorithm can just query elements at random until a pair $x \neq y$ is found such that $f(x) = f(y)$ and then output $a = x \oplus y$. The birthday paradox tells us that the expected number of queries until a collision is found is of order $\Omega(2^{n/2})$.

Let us observe that the periodicity of f induces a partition of the 2^n input strings into two sets X and $\overline{X} = \{x \oplus a : x \in X\}$ each with cardinal 2^{n-1} such that $f(x)$ takes different values for each $x \in X$ (and the same for $\overline{X}$).

If we apply this to the expression $\sum\limits_{x,y\in\mathbb{Z}_2^n} (-1)^{x\cdot y}$ we can identify each term in X with its counterpart in $\overline{X}$ and rewrite it as $\sum\limits_{x\in X, y\in\mathbb{Z}_2^n} ((-1)^{x\cdot y} + (-1)^{(x\oplus a)\cdot y})$, which equals $\sum\limits_{\substack{x\in X, y\in\mathbb{Z}_2^n \\ y\cdot a=0}} 2\cdot(-1)^{x\cdot y}$.

Thus, the state $|\widehat{f}\rangle$ obtained by quantum Fourier sampling is

$$|\widehat{f}\rangle = \frac{1}{2^n} \sum_{x,y \in \mathbb{Z}_2^n} (-1)^{x \cdot y} |y\rangle |f(x)\rangle = \frac{1}{2^{n-1}} \sum_{\substack{x \in X, y \in \mathbb{Z}_2^n \\ y \cdot a = 0}} (-1)^{x \cdot y} |y\rangle |f(x)\rangle.$$

Performing a measurement of the first register in the computational basis will lead to a random result $y_i = y \in \mathbb{Z}_2^n$ such that $y_i \cdot a = 0$. Thus, we gain *one bit* of information about the period a. We do this repeatedly and obtain a set of equations.

Now let us take a deeper look at the meaning of the outcomes y_i. These y_i form a subspace of the n-dimensional $\mathbb{Z}_2$-vector space of all n-bit strings. A solution $a \neq 0$ is completely determined if among the y_i there are $n-1$ linearly independent vectors.
If we have a set of y_i that does not yet span a space of dimension $n-1$, then it contains $\leq 2^{n-2}$ out of the 2^{n-1} possible values for y. Thus, the probability that the next y will fall outside the current space is $\geq \frac{1}{2}$.

Hence, after $O(n)$ repetitions, the algorithm will have enough information to determine a with probability exponentially close to 1. Fig. 6.5 shows the circuit that implements Simon's algorithm.

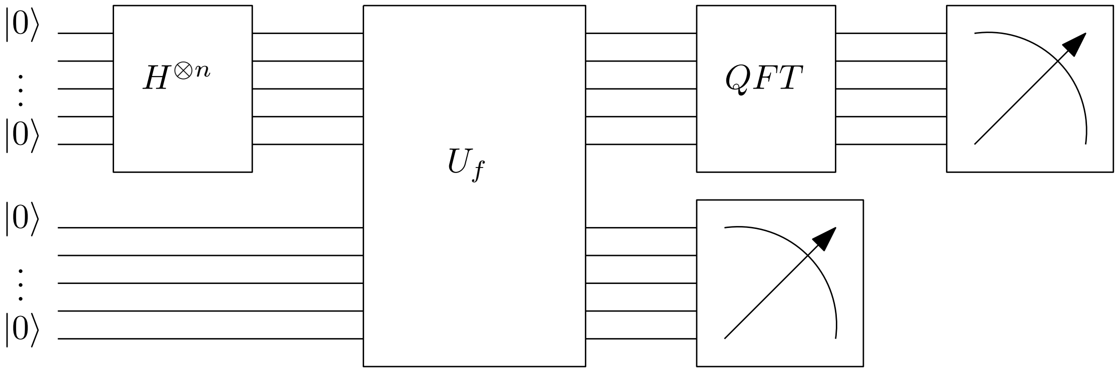

FIG. 6.5. Simon's Algorithm. In this particular case, $QFT = H^{\otimes n}$.

6.4.3. Grover's Algorithm. Unstructured search is one of the most encountered fundamental problems one can find in information processing. We can model it as a database that contains N items, one of which has a special mark, given by a function $f : \{1, \ldots, N\} \longrightarrow \{0, 1\}$ such that $\#f^{-1}(\{1\}) = 1$. If we wish to find the marked item and the database has no special structure (e.g., not sorted), any deterministic or randomized classical algorithm that succeeds to solve the problem will have to make $\Omega(N)$ queries to f on average, since any query has $1/N$ success probability and otherwise discloses no information about the location of the marked item.

The core of the algorithm is the repeated application of the *diffusion transform* D, which is an inversion about the mean (the average value) as can be seen in Fig. We consider an arbitrary superposition $|\psi\rangle = \sum_{y \in \{0,1\}^n} \alpha_y |y\rangle$ on n qubits, with $\alpha_y \in \mathbb{C}$, $\sum_y |\alpha_y|^2 = 1$ and we put $\alpha = \frac{1}{2^n} \sum_y \alpha_y$ be the average of the amplitudes of all basis states in $|\psi\rangle$. The diffusion transform D is such that it maps the amplitude of a basis state to its symmetric respect to α, i.e., $\alpha_y \mapsto 2\alpha - \alpha_y$.

The implementation of D can be done with $D = H^{\otimes n} R H^{\otimes n}$, where R is the diagonal unitary matrix which flips the sign of all basis states except $|0^{\otimes n}\rangle$. [**32**] shows the implementation of R with elementary gates. We see that this actually works, since

$$
\begin{aligned}
H^{\otimes n} R H^{\otimes n} |\psi\rangle &= \frac{1}{\sqrt{2^n}} H^{\otimes n} R \sum_{x,y} (-1)^{x\cdot y} \alpha_y |x\rangle \\
= -\frac{1}{\sqrt{2^n}} H^{\otimes n} &\left(\sum_{x,y} (-1)^{x\cdot y} \alpha_y |x\rangle \right) + \frac{2}{\sqrt{2^n}} \left(\sum_y \alpha_y \right) H^{\otimes n} |0^{\otimes n}\rangle \\
&= -\sum_y \alpha_y |y\rangle + 2\alpha \sum_x |x\rangle = \sum_y (2\alpha - \alpha_y)|y\rangle.
\end{aligned}
$$

On the other hand, the gate $S_f = (\mathbb{I}^{\otimes n} \otimes H) \cdot U_f \cdot (\mathbb{I}^{\otimes n} \otimes H)$ can introduce a phase $-1 = e^{i\pi}$ on every state $|y\rangle$ such that $f(y) = 1$ by using an ancillary qubit initialized to $|1\rangle$ in the last register, as seen in Fig. 6.6.

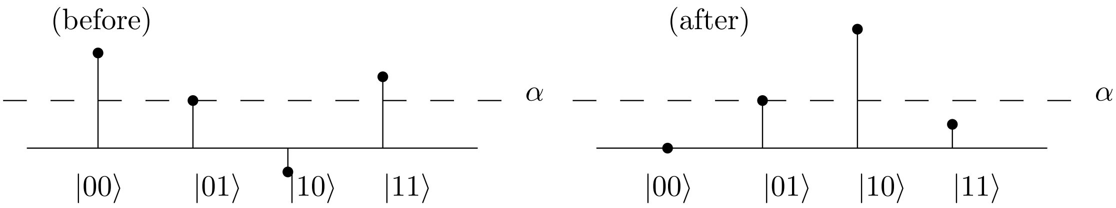

FIG. 6.6. Idea of the Diffusion Transform D: an inversion about the mean or the average α.

Now let us see how to use this to solve our problem: Suppose we first prepare a uniform superposition over all $N = 2^n$ basis states on n qubits. One application of S_f will change the sign of only the basis state which corresponds to the marked item to its opposite.
So, all unmarked items will have amplitude $\frac{1}{\sqrt{N}}$ whereas the marked item $-\frac{1}{\sqrt{N}}$ on first iteration; moreover, $\alpha^{(1)}$ will be close to $\frac{1}{\sqrt{N}}$.

Now, applying the transformation D, all unmarked states will be close to average, so will roughly remain the same. However, the marked item amplitude will go from $-\frac{1}{\sqrt{N}}$ to $\frac{2}{\sqrt{N}} - (-\frac{1}{\sqrt{N}}) = \frac{3}{\sqrt{N}}$.

Repeating this process $c\sqrt{N}$ times, for some constant c, it will lead the marked item's amplitude to be close to 1 and thus, a measurement in the computational basis will give the marked item with high probability.

Grover's algorithm is represented in Fig. 6.7.

Remark 6.9. The exact number of inner loops is important since the amplitudes of the unmarked states evolve as $\alpha^{(i)} = \frac{1}{\sqrt{N-1}} \cos((2i+1)\theta)$ and the amplitudes of the marked state as $\beta^{(i)} = \sin((2i+1)\theta)$, where θ is such that $N \sin^2\theta = 1$, i.e., they oscillate [**32**].

6.4.4. Shor's Algorithm. The problem of factoring large integers is believed to be hard and thus it is the heart of many cryptographic systems which are used today. Factoring is in NP and the current best known classical algorithms run in sub-exponential time in the number of bits in the input.
Peter Shor made an important breakthrough in 1994 [**60**] when he published a polynomial-time

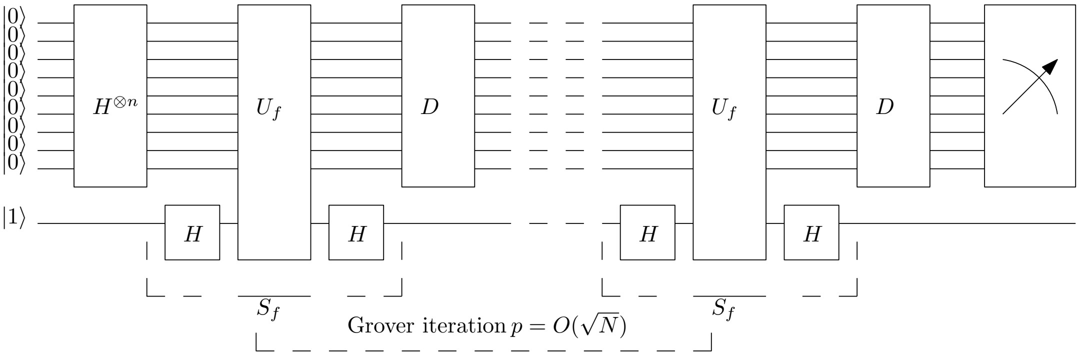

FIG. 6.7. The circuit for Grover's Algorithm.

quantum algorithm for factoring. Its core is similar to Simon's period finding algorithm, but over another group [**62**].

Remark 6.10. We shall outline a classical polynomial-time reduction from factoring to period-finding first.

Let $N \in \mathbb{Z}$ be the integer we wish to factor, i.e., we aim to find $q|N$ a non-trivial factor of N. Let us assume that we are given x such that $x \neq \pm 1 \mod N$ and $x^2 = 1 \mod N$. The latter can be written as $N|(x^2-1) = (x-1)(x+1)$. This means that N has a common factor with $x+1$ or $x-1$, which are different from N by hypothesis. In $O(\log N)$ one can compute this common factor $\gcd(N, x \pm 1)$ using Euclid's algorithm.

In order to compute such an x, let us take an integer $y \in \mathbb{Z}_N$. We can assume it is coprime with N, otherwise we already found a factor using Euclid's algorithm. Consider the function

$$f: \begin{array}{ccc} \mathbb{Z} & \longrightarrow & \mathbb{Z}_N \\ x & \longmapsto & y^x \end{array},$$

which is periodic with period $a = ord(y)$, the order of y (the smallest positive integer such that $y^a = 1 \mod N$).

It can be shown via the Chinese remainder theorem that a y taken at random from $\{0, \ldots, N-1\}$ will have even order $\mod N$ and it will be such that $y^{a/2} \neq \pm 1 \mod N$ with probability at least a constant [**50**].

Remark 6.11. There are two exceptions to this result: if $N = 0 \mod 2$ or if $N = p^s$ for some prime number p, which can be tested beforehand with efficient classical algorithms; so they can be ruled out.

So it suffices to pick y at random, compute its order and if its even, compute $\gcd(y^{a/2} \pm 1, N)$ which will give a non trivial factor of N.

There still remains a loose end, which is the computation of $a = ord(y)$.

Let us state the problem:
Input: An integer $N \in \mathbb{N}^*$ and a function $f : \mathbb{Z} \longrightarrow X$, where X is a set.

Assumption: $\exists a < N$ such that f is a-periodic:

$$\forall x, y \in \mathbb{Z}, \quad x = y \mod a \Rightarrow f(x) = f(y).$$

Output: a.

Remark 6.12. In the particular case where $f : \mathbb{Z}_M \longrightarrow X$ is a-periodic with $a|M$, we have $|\hat{f}\rangle = \frac{1}{M} \sum_{x,y \in \mathbb{Z}_M} \omega^{x \cdot y} |y\rangle |f(x)\rangle = \frac{1}{M} \sum_{x \in \mathbb{Z}_a, y \in \mathbb{Z}_M} \left(\sum_{j=0}^{M/a-1} \omega^{(x+ja)y} \right) |y\rangle |f(x)\rangle$. When $ya = 0 \mod M$ ($ay = cM$ for some c), we have a geometric sum which gives $\sum_{j=0}^{M/a-1} \omega^{(x+ja)y} = \frac{M}{a} \omega^{xy}$ and gives 0 otherwise. So, the state $|\hat{f}\rangle$ is

$$|\hat{f}\rangle = \frac{1}{a} \sum_{x \in \mathbb{Z}_a} \left(\sum_{c \in \mathbb{Z}_a} \omega^{xcM/a} |cM/a\rangle \right) |f(x)\rangle. \tag{6.1}$$

A measurement on the first register will give a random $y = cM/a$ with uniform probability. y and M are known and we need to solve $y/M = c/a$ in order to get a. At most $\lceil \log_2 a \rceil$ elements of $\mathbb{Z}_a$ are not co-prime with a (to see this, decompose a into product of primes). Since c is taken uniformly at random, after $\lceil \log_2 a \rceil = O(\log N)$ repetitions we obtain with high probability a (we express y/M as a reduced fraction and take its denominator, which gives a).

In the general case $f : \mathbb{Z} \longrightarrow X$, two difficulties arise: Firstly, the domain of f must be cut off. To do this, we take an integer $M > N$ and restrict $f_M = f|_{\mathbb{Z}_M}$. To guarantee enough periodicity, we have to pick M large enough. The second difficulty is that a is unknown so we cannot guarantee M to be an exact multiple of a. Hence, f_M will not be perfectly periodic and this will lead to some extra terms in the superposition (6.1). Luckily for us, they will have small norm in comparison to the relevant data.

More precisely, [**32**] shows that there is a constant probability of measuring a y satisfying

$$\left| \frac{y}{M} - \frac{c}{a} \right| \leq \frac{1}{2M}, \quad c \in \{0, \dots, a-1\}.$$

But two distinct fractions with denominator $\leq N$ must be at distance $\geq \frac{1}{N^2}$ so it suffices to choose $M > N^2$ to guarantee that $\frac{c}{a}$ is the unique fraction with denominator $\leq N$ at distance $\leq \frac{1}{2M}$ from $\frac{y}{M}$. This can be found by continued fraction expansion and the total run-time of the algorithm is still $O(poly(\log N))$.

6.4.5. Phase Estimation Algorithm.

Let U be a computation of order n and let $|u\rangle \in \mathcal{H}$ be an eigenvector of U. The corresponding eigenvalue can be written in the form $e^{2\pi i \varphi}$. The problem can be presented as follows:
Input: U and $|u\rangle$.
Output: r bits $\{\varphi_1, \dots, \varphi_r\}$ of the binary expansion $\varphi = 0.\varphi_1\varphi_2\cdots$.

The algorithm that solves this problem is known as Kitaev Algorithm [**34, 35**]. It is depicted in Fig. 6.8 and its steps are the following:

1. We initialize the state in $\mathcal{H}^{\otimes m} \otimes \mathcal{H}^{\otimes n}$ in the value $|0^{\otimes m}\rangle |u\rangle$.
2. We apply a Hadamard gate to each of the first m qubits and obtain

$$(H^{\otimes m} \otimes \mathbb{I}_{2^n}) |0^{\otimes m}\rangle |u\rangle = |h_m\rangle |u\rangle.$$

3. We perform a series of controlled q-computations $C_{m-l+1}(U^{2^{l-1}})$ for $l \in 1 \dots m$ defined as follows:

$$C_{m-l+1}(U^{2^{l-1}})(|\varphi_1 \cdots \varphi_m\rangle|u\rangle) = \begin{cases} |\varphi_1 \cdots \varphi_m\rangle|u\rangle & \text{if} \quad \varphi_{m-l+1} = 0 \\ |\varphi_1 \cdots \varphi_m\rangle(U^{2^{l-1}}|u\rangle) & \text{if} \quad \varphi_{m-l+1} = 1 \end{cases}$$

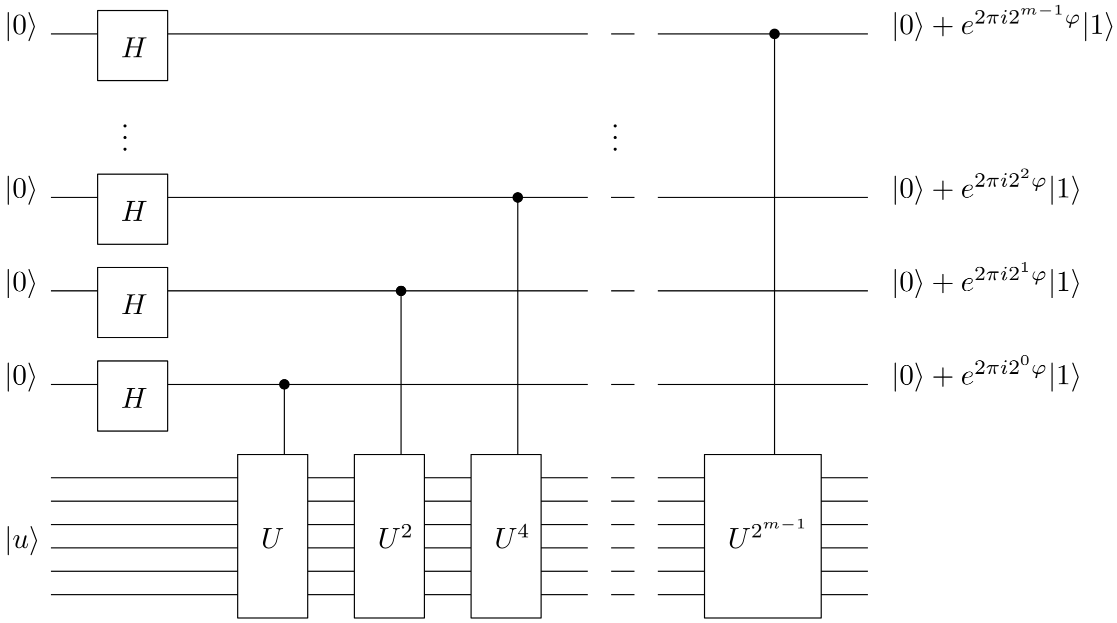
$|0\rangle$
H
$|0\rangle + e^{2\pi i 2^{m-1}\varphi}|1\rangle$
$|0\rangle$
H
$|0\rangle + e^{2\pi i 2^{2}\varphi}|1\rangle$
$|0\rangle$
H
$|0\rangle + e^{2\pi i 2^{1}\varphi}|1\rangle$
$|0\rangle$
H
$|0\rangle + e^{2\pi i 2^{0}\varphi}|1\rangle$
$|u\rangle$
U
U^2
U^4
$U^{2^{m-1}}$

FIG. 6.8. Steps 1-3 of Kitaev's Algorithm for phase estimation.

Observe that the action of $U^{2^{l-1}}$ only changes $|u\rangle$ by a factor which is either 1 or $e^{2\pi i 2^{l-1}\varphi}$ (depending on the controlling qubit). We can move this factor next to the controlling qubit and, thus, we have the state in the first register at the end of this step which is

$$\frac{1}{\sqrt{2^m}}(|0\rangle + e^{2\pi i 2^{m-1}\varphi}|1\rangle)(|0\rangle + e^{2\pi i 2^{m-2}\varphi}|1\rangle) \cdots (|0\rangle + e^{2\pi i 2^{0}\varphi}|1\rangle).$$

As $e^{2\pi i k} = 1 \forall k \in \mathbb{Z}$, we may rewrite this expression using the binary expansion of φ:

$$\frac{1}{\sqrt{2^m}} \bigotimes_{l=1}^{m} (|0\rangle + e^{2\pi i 0.\varphi_{m-l+1}\cdots\varphi_m}|1\rangle).$$

Note that, according to example this is exactly $QFT_{\mathbb{Z}_{2^m}}|\varphi\rangle$.

4. We perform the QFT of the first m qubits in reverse order to undo this operation.
5. We measure the first register onto the computational basis.

Remark 6.13. Observe that Kitaev's algorithm supplies φ exactly in the case when it can be expressed using m bits.

If this is not the case, when we apply $QFT^{\dagger}_{\mathbb{Z}_{2^m}} \otimes \mathbb{I}_{2^n}$ we do not obtain $|\varphi\rangle|u\rangle$, but a superposition of the form $\sum_l a_l|l\rangle|u\rangle$, a difficulty which can be overcome [**50**] in order to obtain the first r bits of φ correctly with good probability $(1-\varepsilon)$ if $m \geq r + \log_2\left(2 + \frac{1}{2\varepsilon}\right)$.

Remark 6.14. As $|u\rangle$ is an eigenvector of U, observe that the state remains intact at the end of the computation even having passed through various U gates.

6.5. Hidden Subgroup Problem

Simon and Shor's algorithms can be seen as particular cases for solving a more general problem, which is called the Hidden Subgroup Problem (HSP). The HSP for Abelian groups can be solved efficiently by a quantum algorithm. Inside this more general frame, one finds several known problems; e.g., the discrete logarithm, finding solutions to Pell's equation or the hidden translation problem: $f, g : \mathbb{Z}_p^n \longrightarrow X$ such that $f(x) = g(x+t)$ for some t we want to find, etc.

More formally, the problem we want to treat it the following:
Input: A finite group G, a set S, and a function $f : G \longrightarrow S$, which is given as a black box.
Assumptions: f is constant on (left) cosets of some unknown subgroup $H \leq G$, and f takes different values on different cosets.
Output: H (a set of generators).

The algorithm follows two basic steps

1. *Preparation of a random coset state.* We prepare the state

$$|\hat{f}\rangle = \frac{1}{|G|}\sum_{x,y\in G}\chi_y(x)|y\rangle|f(x)\rangle = \frac{1}{|G|}\sum_{\substack{c\in G/H\\ y\in G}}\left(\sum_{x\in H}\chi_y(cx)\right)|y\rangle|f(cH)\rangle.$$

   We measure the last register. Thus, we select a random coset cH of $H \leq G$ and project the state onto

$$|\hat{f}'\rangle = \frac{1}{\sqrt{|G|\cdot|H|}}\sum_{y\in G}\left(\sum_{x\in H}\chi_y(cx)\right)|y\rangle.$$

2. *Fourier sampling.* We measure the first register and we get $y \in G$ with probability

$$\left|\sum_{x\in H}\chi_y(cx)\right|^2 = \left|\chi_y(c)\sum_{x\in H}\chi_y(x)\right|^2 = \left|\sum_{x\in H}\chi_y(x)\right|^2.$$

Remark 6.15. On step 2, we get a random $y \in H^\perp = \{y \in G, \chi_y(h) = 1\ \forall h \in H\}$, since $|\hat{f}'\rangle$ is a uniform superposition over all $y \in G$ such that $\chi_y(h) = 1\forall h \in H$; [**32**]. Such y can be viewed as a linear constraint on H. Thus, with $O(\log|H|)$ independent constraints we can effectively reconstruct H. This is, by quantum Fourier Sampling a sufficient number of times, we find a generating set for H.

6.6. General overview

In order to give a general view of which are the most remarkable quantum algorithms known today −which present a non-trivial speedup with respect to the best known classical counterpart−, in this section we give list them, together with the kind of speedup they offer.

We have classified the algorithms into three classes: those which solve algebraic and number theory problems, those which are based on an oracle function (like Grover's U_f function), and approximation algorithms.

SP stands for Superpolynomial, P for Polynomial, C for Constant and E for Exponential.

Since the recent boost of activity in the field of quantum computation has given it a big growth, the current list of algorithms is quite large. We recommend [**64**] for an up-to-date listing of algorithms, as well as the formal explanation of all the problems listed below, with respective references.

Algebraic and Number Theory problems:

Algorithm	Speedup
Integer factorization (Shor)	SP
Discrete logarithm	SP
Pell's Equation	SP
Principal Ideal	SP
Unit Group	SP
Gauss Sums	SP
Exponential congruences	P
Matrix Elements of Group Representations	SP

Oracle-Based problems:

Algorithm	Speedup
Unstructured Search (Grover)	P
Abelian Hidden Subgroup	SP
Non-Abelian Hidden Subgroup	SP
Bernstein-Vazirani	P
Deutsch-Josza	E
NAND-Tree	P
Gradients and Quadratic Forms	P
Hidden Shift	SP
Linear Systems	SP
Ordered Search	C
Graph Properties (Adjacency Matrix representation)	P
Graph Properties (Adjacency List representation)	P
Welded Tree	SP
Collision Finding	P
Matrix Commutativity	P
Group Commutativity	P
Hidden Nonlinear Structures	SP
Center of Radial Function	P
Group Order and Membership	SP
Group Isomorphism	SP
Statistical Difference	P
Finite Rings and Ideals	SP
Counterfeit Coins	P

Approximation problems:

Algorithm	Speedup
Quantum Simulation	SP
Knot Invariants	SP
Three-Manifold Invariants	SP
Partition Functions	SP
Adiabatic Optimization	(Unknown)
Zeta Functions	SP
Weight Enumerators	SP
Simulated Annealing	P
String Rewriting	SP
Matrix Powers	SP
Matrix Product Verification	P

Chapter 7
Quantum Entropy and Information

In this chapter, we present the generalization of the concept of entropy to a quantum level, as well as Schumacher's theorem of data compression and the Holevo bound for accessible information. We follow the exposition one can find in [**17**], along with some useful comments from [**48**]. We introduce several important quantum channels and discuss its capacities, as well as how they interact with the qubits that pass through them. We conclude with the presentation of quantum entanglement in a formal manner, discussing its methods for detection in the general case.

7.1. Quantum Entropy

Definition 7.1. The quantum entropy $S(A)$ of a system A with density matrix ρ_A is defined as

$$S(A) = -Tr(\rho_A \log \rho_A).$$

The quantum joint entropy $S(A, B)$ of a composite system with two components A and B is defined as

$$S(A, B) = -Tr(\rho_{AB} \log \rho_{AB}),$$

where ρ_{AB} is the density matrix of the composite system AB.
The quantum conditional entropy $S(A|B)$ is defined as

$$S(A|B) = S(A, B) - S(B).$$

The quantum mutual information of two subsystems A and B of a composite system AB is defined as

$$S(A : B) = S(A) + S(B) - S(A, B) = S(A) - S(A|B) = S(B) - S(B|A).$$

Remark 7.1. If λ_x are the eigenvalues of ρ, then von Neumann's definition of entropy can be rewritten [**48**] as

$$S(\rho) = -\sum_x \lambda_x \log \lambda_x,$$

where logarithms are taken in base 2 and the same conventions of Remark 2.3 apply.

Remark 7.2. In fact, this is what motivated Definition 7.1, and it agrees with what we defined in Section 2.1.

Remark 7.3. For classical random variables X and Y, one has $H(X) \leq H(X, Y)$ [**17**]. Its intuitive explanation is that the uncertainty of a random variable X cannot be more than the uncertainty of the pair of random variables X and Y.

However, this in general is false in the quantum picture. Indeed, for a bipartite quantum system AB, $S(A)$ can exceed $S(A,B)$: In the case of a maximally entangled state of two d-dimensional systems $S(A,B) = 0$, since the joint state is a pure state; however $\rho_A = Tr_B \rho_{AB} = \frac{1}{d}\mathbb{I}_A$ and $S(A) = \log d > 0$. We should also observe that, in such a case, the quantum conditional entropy is negative $S(B|A) = S(A,B) - S(A) = -S(A) < 0$ (see also Remark 7.16).

Properties 7.1. The von Neumann entropy is strongly subadditive [**17**]:

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}), \tag{7.1}$$

for a tripartite system.

This has several important consequences:

1. *Conditioning reduces entropy:* When conditioning on two systems B and C, the entropy is less than conditioning on just system B:

   $$S(A|BC) \leq S(A|B).$$

   Proof: Since $S(A|BC) = S(A,B,C) - S(B,C)$, where the terms on the right hand side are the von Neumann entropies of their respective systems, it follows from (7.1) that $S(A|BC) \leq S(A,B) + S(B,C) - S(B) - S(B,C) = S(A|B)$. □
2. *Discarding quantum systems never increases mutual information:* i.e.,

   $$S(A:B) \leq S(A:BC).$$

   Proof: $S(A:BC) = S(A) + S(B,C) - S(A,B,C)$ which, by (7.1), is $\geq S(A) + S(B,C) - S(A,B) - S(B,C) + S(B) = S(A:B)$. □
3. *Quantum operations never increase mutual information:* Given a composite quantum system AB and Φ a CPT (Definition 7.2) map acting on the subsystem B alone. Let us denote $A'B'$ the composite system after this action. Then

   $$S(A':B') \leq S(A:B).$$

   The proof of this point can be found at [**17**].

Definition 7.2. A quantum operation, or superoperator Φ is a linear, completely positive, trace-preserving map (CPT) which take density matrices to density matrices. More precisely,

$$\Phi : \rho \longmapsto \rho', \quad \rho \in \mathcal{B}(\mathcal{H}), \quad \rho' \in \mathcal{B}(\mathcal{H}'), \quad \rho, \rho' \geq 0, \quad Tr\rho = Tr\rho' = 1.$$

A quantum operation captures the dynamical change to the state of a system, which occurs as the result of some physical process: time evolution of an open system, compression data from a quantum operation source and transmission of quantum information through a noisy quantum channel are examples of this. In the latter case, $\mathcal{H}$ and $\mathcal{H}'$ are referred to as the input and output Hilbert spaces of the channel Φ.

7.2. Data Compression

Definition 7.3. A quantum information source is defined by a set of pure state $\{|\psi_k\rangle\}_k$, with corresponding probabilities $\{p_k\}_k$. Each of these states acts on a Hilbert space $\mathcal{H}$. From the information theory point of view, the $|\psi_k\rangle$ are signals produced with the source, each with probability p_k. Thus, a quantum information source can be characterized as

$$\{\rho, \mathcal{H}\},$$

where $\rho = \sum\limits_k p_k|\psi_k\rangle\langle\psi_k|$ is a density matrix acting on $\mathcal{H}$.

Remark 7.4. The states $\{|\psi_k\rangle\}_k$ need not be mutually orthogonal.

In classical information theory, data compression corresponds to a reduction in the number of bits required to store the information emitted by a classical information source. In the quantum case, the idea behind data compression is similar, replacing bits by qubits and the nature of the source (classical for quantum). The quantity we aim to compress, in this case, is the *dimension of the Hilbert space* $\widetilde{\mathcal{H}_n}$:

More precisely, we consider sequences of density matrices $\rho^{(n)}$, for arbitrary n, acting on Hilbert spaces $\mathcal{H}_n$ of increasing dimension N_n with n:

$$\rho^{(n)} = \sum_k p_k^{(n)}|\psi_k^{(n)}\rangle\langle\psi_k^{(n)}|, \qquad p_k^{(n)} \geq 0, \quad \sum_k p_k^{(n)} = 1.$$

In order to compress data from such a source, we encode each signal state $|\psi_k^{(n)}\rangle$ by a state $\widetilde{\rho}_k^{(n)} \in \mathcal{B}(\widetilde{\mathcal{H}_n})$. We set $d_c(n) = \dim\widetilde{\mathcal{H}_n}$ and we require $d_c(n) \leq N_n$ $\forall n$.

Definition 7.4. A compression scheme is a CPT map

$$\mathcal{C}^{(n)} : |\psi_k^{(n)}\rangle\langle\psi_k^{(n)}| \longmapsto \widetilde{\rho}_k^{(n)} \in \mathcal{B}(\widetilde{\mathcal{H}_n}).$$

And a corresponding decompression scheme is a CPT map

$$\mathcal{D}^{(n)} : \mathcal{B}(\widetilde{\mathcal{H}_n}) \longrightarrow \mathcal{B}(\mathcal{H}_n).$$

Our goal will be to minimize $d_c(n)$ while preserving the existence of a reliable compression $-$ decompression scheme.

Definition 7.5. We define the rate of compression as

$$R_n := \frac{\log\dim\widetilde{\mathcal{H}_n}}{\log\dim\mathcal{H}_n} = \frac{\log d_c(n)}{\log N_n}.$$

The usual case is when $\mathcal{H}_n$ corresponds to the Hilbert space of n qubits, so $N_n = 2^n$; hence, $\log N_n = n$.

We are interested in finding the optimal rate of data compression, which in this case is given by

$$R_\infty := \lim_{n\to\infty} R_n = \lim_{n\to\infty} \frac{\log d_c(n)}{n}.$$

Remark 7.5. Since we do not require the states $\{|\psi_k^{(n)}\rangle\}_k$ be orthogonal, a new problem arises: they are, in general, not completely distinguishable.

If we wished to reconstruct perfectly a quantum signal state from its compressed version, it would turn out to be a too stringent task because of this reason.

Instead, a reasonable requirement for the reliability of the compression-decompression scheme is that a nearly indistinguishable state from the original one can be reconstructed from $\widetilde{\rho}_k^{(n)}$.

Definition 7.6. A measure of indistinguishability which is useful for this purpose is the average ensemble fidelity, which is defined as

$$F_n := \sum_k p_k^{(n)}\langle\psi_k^{(n)}|\mathcal{D}^{(n)}(\widetilde{\rho}_k^{(n)})|\psi_k^{(n)}\rangle.$$

Fidelity satisfies [**17, 51**]

$$0 \leq F_n \leq 1 \quad \text{and} \quad F_n = 1 \iff \mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)}) = |\psi_k^{(n)}\rangle\langle\psi_k^{(n)}| \ \forall k.$$

Definition 7.7. A compression-decompression scheme is said to be reliable if

$$\lim_{n\to\infty} F_n = 1.$$

Remark 7.6. The idea which leads to data compression is that some signal states occur with a higher probability than others do. In the quantum case this is translated into the following: such signals span a subspace of the original Hilbert space of the source, which is referred to as the typical subspace.

7.2.1. Schumacher's Theorem for Memoryless Quantum Sources. When considering a memoryless quantum source, its density matrix ρ_n acts on the tensor product Hilbert space $\mathcal{H}^{\otimes n}$ and is given by the tensor product $\pi^{\otimes n}$, where π is the density matrix of a single qubit. We announce Schumacher's quantum coding theorem for i.i.d. quantum sources, the proof of which can be found at [**17**].

Theorem 7.1. (Schumacher's quantum coding theorem)
Let $\{\rho_n, \mathcal{H}_n\}$ be an i.i.d. quantum source: $\rho_n = \pi^{\otimes n}$ and $\mathcal{H}_n = \mathcal{H}^{\otimes n}$. If $R > S(\pi)$, then there exists a reliable compression scheme of rate R. If $R < S(\pi)$ then any compression scheme rate of R is unreliable.

7.3. Quantum Channels

For the following examples we will need these results [**51**]:

Theorem 7.2. Any state transformation $\rho \mapsto \mathcal{E}(\rho)$ can be written in the form called operator sum representation:

$$\mathcal{E}(\rho) = \sum_p A_p \rho A_p^\dagger,$$

where the operation elements A_p satisfy $\sum\limits_p A_p A_p^\dagger = \mathbb{I}$.

Conversely, all linear mappings of this form are state transformations.

Definition 7.8. $\mathcal{E}$ is completely positive if $\mathcal{E} \otimes id_n$ preserves positivity, for the identical mapping $id_n : M_n(\mathbb{C}\ ) \longrightarrow M_n(\mathbb{C}\ )$ on any matrix algebra.

This result [**51**] was first proved by Kraus:

Theorem 7.3. Let $\mathcal{E} : M_n(\mathbb{C}\ ) \longrightarrow M_k(\mathbb{C}\ )$ be a linear map. $\mathcal{E}$ is completely positive if, and only if, it admits a representation

$$\mathcal{E}(A) = \sum_u V_u A V_u^\dagger$$

by means of some linear operators $V_u : \mathbb{C}^{\ n} \longrightarrow \mathbb{C}^{\ k}$.
This is also equivalent to: the representing block matrix

$$(X_{ij})_{1 \leq i,j \leq k} \in M_k(\mathbb{C}\ ) \otimes M_n(\mathbb{C}\ )$$

is positive.

Remark 7.7. This operator sum representation is called Kraus representation and it is not unique.

Example 7.1. *Bit flip channel.* This is a single qubit channel which flips the qubit sent through it with probability $(1-p)$. If ρ is the input state to the channel, then the output is

$$\Phi(\rho) = p\sigma_x\rho\sigma_x + (1-p)\rho.$$

The corresponding Kraus operators are $A_1 = \sqrt{p}\sigma_x$ and $A_2 = \sqrt{p-1}\mathbb{I}$.

Example 7.2. *Depolarizing channel.* We present two different approaches for the analysis of this channel:

- This channel leaves intact the input qubit with probability $1-p$, and it introduces the following errors, each with probability $p/3$: bit flip (σ_x), phase flip (σ_z) and combined flip (σ_y). Hence, the output of the channel is

  $$\Phi(\rho) = (1-p)\rho + \frac{p}{3}\left(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z\right).$$

  In such case, we have four Kraus operators:

  $$A_1 = \sqrt{1-p}\mathbb{I}, \quad A_2 = \sqrt{\frac{p}{3}}\sigma_x, \quad A_3 = \sqrt{\frac{p}{3}}\sigma_y, \quad A_4 = \sqrt{\frac{p}{3}}\sigma_z.$$

- We can also think of the depolarizing channel to leave a qubit unaffected with a certain probability $1-q$ and to replace its state with the *completely mixed state* $\frac{1}{2}\mathbb{I}$ with probability q (the center of the Bloch Sphere):

  $$\Phi(\rho) = (1-q)\rho + \frac{q}{2}\mathbb{I}.$$

  Observe that, since $\rho + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z = 2\cdot\mathbb{I}\ \forall\rho\in\mathcal{B}(\mathcal{H})$,

  $$\Phi(\rho) = \left(1 - \frac{3}{4}q\right)\rho + \frac{q}{4}\left(\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z\right),$$

  which means that $p = \frac{3}{4}q$.
  Since the state $\frac{\mathbb{I}}{d}$ represents the completely mixed state in a d-dimensional Hilbert space, the depolarizing channel can be generalized to

  $$\Phi(\rho) = (1-p)\rho + \frac{p}{d}\mathbb{I}.$$

Example 7.3. *Amplitude damping channel.* This channel is a model for energy dissipation. We have a 2-level atom in the excited state $|1\rangle$, which has probability p to decay to its ground state $|0\rangle$ due to spontaneous emission of a photon.

This can be decomposed into the Kraus operators

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}.$$

Let us analyze how they interact with each state:

$$A_1|0\rangle = |0\rangle, \quad A_1|1\rangle = \sqrt{1-p}|1\rangle$$
$$A_2|0\rangle = 0, \quad A_2|1\rangle = \sqrt{p}|0\rangle.$$

Operator A_1 describes how the state evolves if there is no decay, whereas operator A_2 describes the decay of the atom from its excited state to the ground state.

So, the action of this channel is

$$\Phi : \left(\begin{array}{cc} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{array} \right) \longmapsto \left(\begin{array}{cc} \rho_{00} + p\rho_{11} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{array} \right).$$

The amplitude damping channel is an example of a quantum channel which takes a *mixed initial state* $\rho = \sum_{i,j} \rho_{ij}|i\rangle\langle j|$ to a pure state $|0\rangle\langle 0|$ asymptotically, since applying the channel n times in succession leads to

$$\lim_{n\to\infty} \phi^n(\rho) = \left(\begin{array}{cc} \rho_{00} + \rho_{11} & 0 \\ 0 & 0 \end{array} \right) = \left(\begin{array}{cc} Tr(\rho) & 0 \\ 0 & 0 \end{array} \right) = |0\rangle\langle 0|.$$

This is intuitively obvious, since an atom in its excited state decays to its ground state eventually.

Example 7.4. *Phase damping channel.* We begin by writing a single-qubit density matrix in its most general form:

$$\rho = \left(\begin{array}{cc} p & \alpha \\ \alpha^* & 1-p \end{array} \right), \quad 0 \le p \le 1,\ \alpha \in \mathbb{C}\ , \quad |\alpha| \le \sqrt{p(1-p)}.$$

The phase damping channel introduces a decay of the off-diagonal terms (which we called decoherences in Remark 3.9). The off-diagonal elements have no classical analogue and the phase damping channel plays a central role in the transition from the quantum to classical world, as we shall see in Section

More precisely, it acts as a rotation (*phase kick*) through an angle θ about the z axis of the Bloch sphere. As seen in section this rotation is described by the matrix

$$R_z(\theta) = \left(\begin{array}{cc} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{array} \right)$$

We shall also assume that the rotation angle is drawn from a Gaussian random distribution $(0, 2\lambda)$:

$$p(\theta) = \frac{1}{\sqrt{4\pi\lambda}} e^{-\frac{\theta^2}{4\lambda}}.$$

The new density matrix ρ' is therefore obtained after averaging over θ:

$$\rho' = \int_{-\infty}^{\infty} p(\theta) R_z(\theta) \rho R_z^\dagger(\theta) d\theta = \left(\begin{array}{cc} p & \alpha e^{-\lambda} \\ \alpha^* e^{-\lambda} & 1-p \end{array} \right).$$

Comparing ρ' with Section 3.3 we see that the phase damping channel maps the coordinates of the Bloch sphere as $x' = e^{-\lambda}x,\ y' = e^{-\lambda}y,\ z' = z$.

Observe that the repeated application of the phase-damping channel leads to coherences dropping to zero exponentially: $\alpha^{(n)} = e^{-\lambda n}\alpha$.

7.4. Accessible Information

Definition 7.9. The maximum information about a random variable X that can be gained through any possible measurement is called accessible information.

$$I_{acc} = \max H(X:Y),$$

where the maximum is taken over all possible measurement schemes.

Remark 7.8. I_{acc} can be viewed as the amount of classical information that can be stored and recovered from a quantum system.

Theorem 7.4. (Holevo bound)
Suppose we have a classical source, characterized by a random variable X which takes values $x \in J = \{1, 2, \dots, M\}$ with probabilities $p(x)$. We encode the symbol x into a quantum state ρ_x and send it to a noiseless quantum channel. Then we perform a measurement on it, described by a finite set of POVM elements $\{E_y\}$. Let Y be the classical random variable corresponding to the outcome of the measurement. Then, for any such measurement that can be done, the mutual information $H(X : Y)$ satisfies the Holevo bound:

$$H(X : Y) \leq \chi,$$

where

$$\chi = S(\rho) - \sum_{x \in J} p_x S(\rho_x),$$

and $\rho = \sum\limits_{x \in J} p_x \rho_x$.

Remark 7.9. The equality is attained if all the ρ_x commute (in that case, they are simultaneously diagonalizable) and the measurement is performed in the simultaneous eigenbasis of all the ρ_x's.

Remark 7.10. The Holevo χ quantity depends on the state ρ, but it also depends on its *preparation*; namely, the ensemble $\mathcal{E} = \{p_x, \rho_x\}$.

When the ensemble consists of pure states, χ reduces to the von Neumann entropy $S(\rho)$. This is because $S(\rho_x) = 0$ if ρ_x is pure.

Notation 7.1. When we want to specify the ensemble from which we derive the Holevo quantity, we shall note it $\chi(\mathcal{E})$.

Properties 7.2. The Holevo bound possesses some remarkable properties [**17**]:

- It is non-negative: $\chi(\mathcal{E}) \geq 0$.
- A quantum operation can never increase the Holevo χ quantity: If $\mathcal{E} = \{p_x, \rho_x\}$ and $\mathcal{E}' = \{p_x, \Phi\rho_x\}$, then
$$\chi(\mathcal{E}') \leq \chi(\mathcal{E}).$$

Remark 7.11. The monotonicity of χ under quantum operations Φ is an indicator that the Holevo bound quantifies the amount of information encoded in a quantum system.
Decoherence, described by an operation Φ, can never increase χ -which is consistent with the intuitive fact that noise never increases information-.

Remark 7.12. In contrast, the von Neumann entropy is not monotonic under quantum operations. We have already seen in Example 7.2 that the depolarizing channel transforms a pure state into a mixed state; thus, increasing the von Neumann entropy S. On the other hand, in Example 7.3, the amplitude damping channel takes a mixed state into a pure state (since the atom ends up decaying to its ground state), thereby reducing the von Neumann entropy. This should not be look at as an information gain, since *every* mixed state decays to the ground state and we lose ability to distinguish between different preparations of the mixed state.

7.5. Quantum Entanglement

In this section, we focus on the study of quantum entanglement, particularizing on the case of 2 particles. We also mention the most used criteria for the detection of entanglement, such as the Peres criterion or the use of entanglement witnesses [**4, 6, 30, 38, 39, 48**].

7.5.1. Formal definition. We consider finite-level systems Σ_i. We denote its set of states by $\mathcal{S}(\Sigma_i)$. We denote its joint system $\Sigma = \sum\limits_i \Sigma_i$.

If we consider a multipartite system consisting of n subsystems, according to classical description, the total state space of the system is the Cartesian product of the n subsystem state spaces. This implies that the total state is always a product state of the n separate subsystems. In contrast, according to quantum formalism, more precisely, Axiom 4, the total Hilbert space H is a tensor product of the subsystems' Hilbert space.
As the dimension of the tensor product space is given by the product of the dimension of each component (whereas the dimension of the Cartesian product space is just the sum) it is clear that, in general, the quantum state of the system cannot be described by the states of all the subsystems separately[1].

Definition 7.10. In the case where a pure state $|\psi\rangle \in \mathcal{S}(\Sigma)$ can be written as the tensor product $|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$, we will say that the state is separable, or a product state. Otherwise, it is entangled.

However, in practice, one finds mixed states, rather than pure. Entanglement of mixed states is no longer equivalent to being non-product, as in the case of Definition 7.10. Instead, the following generalized definition is used [**30**]:

Definition 7.11. A mixed state of n systems $\{\Sigma_i\}_i$ is entangled if it cannot be written as a convex combination of product states

$$\rho \neq \sum_j \lambda_j \rho_1^j \otimes \cdots \otimes \rho_n^j, \quad \rho_i^j \in \mathcal{S}(\Sigma_i), \sum_j \lambda_j = 1, \lambda_j \geq 0.$$

The states which are not entangled according to the definition are called separable.

Remark 7.13. Although one does not find a consensus in the literature, in the pure case, separable states are also called product states, whereas separable is a term left for mixed states.

In short, entanglement is what allows to perform non-classical tasks.

Remark 7.14. In practice, it is hard to decide if a given state is separable or entangled, based on Definition 7.11. The so-called *separability problem* remains still open today and a complete (operative) characterization of entanglement has not yet been given, achieving only partial results [**30, 38, 48**].

In this chapter, we shall present the most relevant of these criteria.

[1] This is what allows to construct exponentially large superpositions with only a linear amount of resources, which is the key to quantum computation.

7.5.2. Entanglement Detection.

Remark 7.15. Unless otherwise stated, we will consider $\Sigma = \Sigma_1 + \Sigma_2$, i.e., the *bipartite* case, until the end of this chapter.

In the case of bipartite pure states, $|\Psi\rangle \in \mathcal{S}(\Sigma)$, the response is elementary: If we express it over a product basis $\{|e_1^i\rangle \otimes |e_2^j\rangle\}_{i,j}$ and we call d_k the dimension of the system Σ_k, we have

$$|\psi\rangle = \sum_{i=0}^{d_1-1} \sum_{j=0}^{d_2-1} A_{ij}^{\Psi} |e_1^i\rangle \otimes |e_2^j\rangle.$$

$|\Psi\rangle$ is a product state if, and only if, the rank of the matrix $A^{\Psi} = (A_{ij}^{\Psi})_{i,j}$ is 1.

In general, the rank $r(\Psi) \equiv r(A^{\Psi}) \leq \min\{d_1, d_2\}$ is called the Schmidt rank of vector $|\Psi\rangle$ and it is equal to either of the ranks of the reduced density matrices

$$\rho_1^{\Psi} = Tr_2(|\Psi\rangle\langle\Psi|), \qquad \rho_2^{\Psi} = Tr_1(|\Psi\rangle\langle\Psi|),$$

which satisfy

$$\rho_1^{\Psi} = A^{\Psi}(A^{\Psi})^{\dagger}, \quad \rho_2^{\Psi} = \left((A^{\Psi})^{\dagger} A^{\Psi}\right)^{T},$$

as we have already seen from Lemma 3.1.

In particular, Theorem 1.5 gives us the Schmidt decomposition: There exists a product bi-orthonormal basis $\{|\tilde{e}_1^i\rangle \otimes |\tilde{e}_2^i\rangle\}$ such that

$$|\Psi\rangle = \sum_{i=0}^{r(\Psi)} \sqrt{p_i} |\tilde{e}_1^i\rangle \otimes |\tilde{e}_2^i\rangle.$$

This argument should make obvious the following proposition:

Proposition 7.1. A bipartite pure state is entangled if, and only if, its Schmidt number is greater than one.

Equivalently, a bipartite pure state is separable if, and only if, the rank of either of the reduced density matrices ρ_1^{Ψ}, ρ_2^{Ψ} is equal to one, or there is a single non-zero Schmidt coefficient $\sqrt{p_i}$.

Remark 7.16. A similar characterization can be made in terms of (quantum conditional) entropy [**17**]:
A pure state $|\psi\rangle$ of a bipartite system $\Sigma_1 + \Sigma_2$ is entangled if, and only if,

$$S(\Sigma_2|\Sigma_1) < 0.$$

However, phenomenons like decoherence (see Section 9.2 or Example 7.2) turn pure states into mixed ones, which may still contain some *"noisy"* entanglement.

According to Definition 7.11, even in the case $n = 2$, it is a hard problem to say if the state is entangled. Its separable decomposition may have nothing to do with its eigenvalue decomposition (as it did in the pure state case). Moreover, many separable states have their eigenvectors entangled [**30**].

Remark 7.17. According to Definition 7.11, the set of separable states, S, is a convex set, and it is invariant under product unitary operations $U_1 \otimes U_2$ [**30**].

7.7.5.2.1. *The PPT criterion.* The characterization of the set of mixed bipartite separable states can be partially made with several entropic necessary conditions [**30**]. However, a much more powerful criterion has been provided by Asher Peres [**6, 30**], which is known as the *positive partial transpose* (PPT) criterion.

Theorem 7.5. (Peres criterion)
A state $\rho \in \mathcal{S}(\Sigma_1 + \Sigma_2)$ is entangled if it does not remain positive under partial transposition:

$$(T \otimes id_2)(\rho) \ngeq 0 \Rightarrow \rho \text{ is entangled.}$$

Here, the operation T is transposition on the indexes corresponding on subsystem Σ_1. The partially transposed state is usually noted ρ^{T_1}, ρ^{T_2} or ρ^{Γ}.

Remark 7.18. The definition of partial transposition is operative, in the sense that fixing some product basis, there is the following correspondence in elements (for example, let us take $\Gamma = T_2$):

$$\langle m|\langle \mu|\rho^{\Gamma}|n\rangle|\nu\rangle \equiv \langle m|\langle \nu|\rho|n\rangle|\mu\rangle.$$

Theorem 7.6. A fundamental fact is [**30, 39**] that the Peres criterion is also sufficient in the cases where the dimension of the corresponding Hilbert spaces are $2 \otimes 2$ or $2 \otimes 3$ $(3 \otimes 2)$.

Definition 7.12. The (entangled) states which are detected with this criterion, i.e., are not positive under partial transposition are called NPT states, whereas the (entangled) states which remain undetected are called PPT (positive under partial transposition).

A representation of this has been depicted in Fig. 7.1.

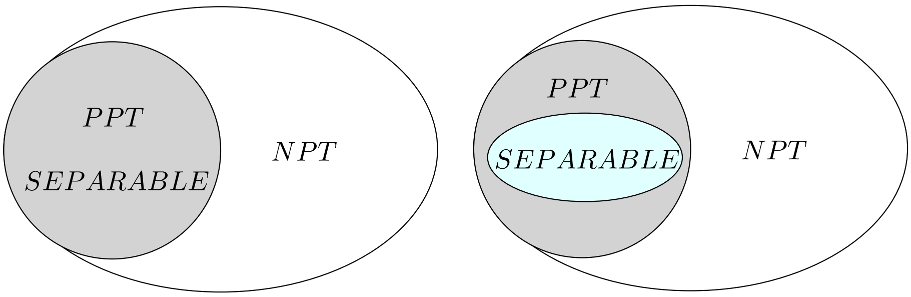

Fig. 7.1. On the left, the classification for $\mathcal{H} = \mathbb{C}^{\,2} \otimes \mathbb{C}^{\,2}$ and $\mathcal{H} = \mathbb{C}^{\,2} \otimes \mathbb{C}^{\,3}$ $(\mathcal{H} = \mathbb{C}^{\,3} \otimes \mathbb{C}^{\,2})$. On the right, the general classification for $\mathcal{H} = \mathbb{C}^{\,m} \otimes \mathbb{C}^{\,n}$, with $m = 2, n \geq 3$ and $\min\{m, n\} \geq 3$.

7.7.5.2.2. *Separability via EW.* Another approach to the separability problem is the consideration of Entanglement Witnesses (EW).

EWs are observables which completely characterize separable states and allow to detect entanglement physically. In Fig. 7.2 we can see its geometric interpretation: We can describe a convex set (see Remark 7.17) by means o hyperplanes.

This suggests the following definition [**13, 38**]:

Definition 7.13. We say that an operator $W = W^{\dagger}$ acting on $\Sigma_1 + \Sigma_2$ is an EW if, and only if,

(i) $\langle e, f|W|e, f\rangle \geq 0$ for all product vectors $|e, f\rangle = |e\rangle \otimes |f\rangle$,

(ii) it has, at least, one negative eigenvalue; i.e., $W \ngeq 0$, and
(iii) $Tr(W) = 1$.

Note that (i) implies that $\langle W\rangle_\rho = Tr(W\rho) \geq 0$, for all ρ separable, whereas (ii) implies that W always detects something and (iii) is nothing but a normalization condition.

We usually denote $\mathcal{D}_{m,n}$ and $\mathcal{D}^{sep}_{m,n}$ the set of density matrices and separable density matrices, respectively, acting on Σ. The set of entangled states detected by W is usually denoted $D_W \equiv \{\rho \in \mathcal{D}_{m,n} : \langle W\rangle_\rho \ngeq 0\}$.

Given two EWs, W_1, W_2, we say that W_1 is finer than W_2 if $D_{W_2} \subseteq D_{W_1}$. Hence, we say that W is optimal if there does not exist any other EW which is finer than W.

This is represented in Fig. 7.2.

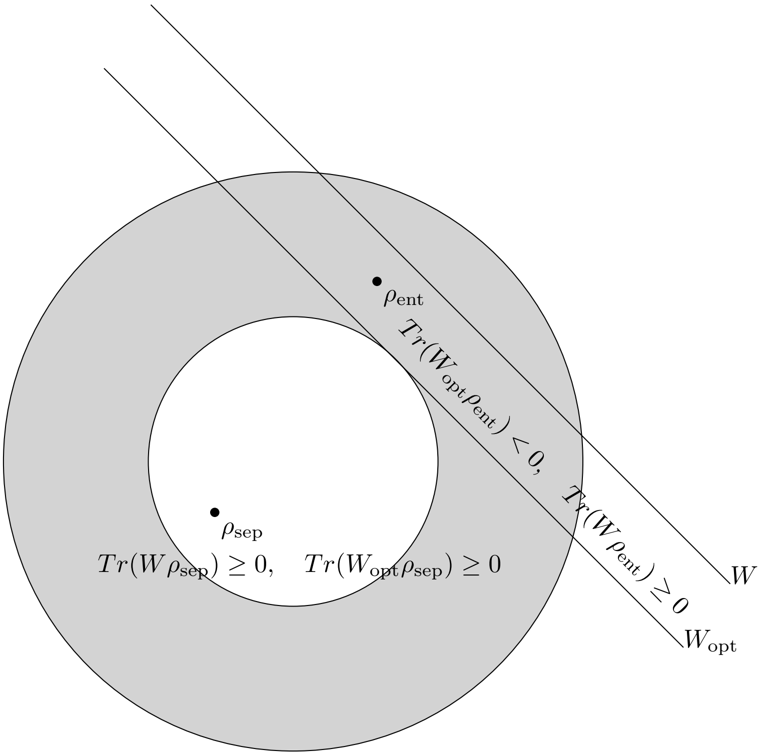
ρ_{ent}
$Tr(W_{\text{opt}}\rho_{\text{ent}}) < 0,$
$Tr(W\rho_{\text{ent}}) \geq 0$
ρ_{sep}
$Tr(W\rho_{\text{sep}}) \geq 0,$
$Tr(W_{\text{opt}}\rho_{\text{sep}}) \geq 0$
W
W_{opt}

FIG. 7.2. The lines represent the hyperplanes which correspond to an entanglement witness W and an optimal entanglement witness W_{opt}. All states not to the right of the hyperplane provide non-negative mean value of the witness, whereas those on the right are detected.

We have the following partition in the set of EWs [**30**]: An EW W is called decomposable (DEW) if it can be written as

$$W = aP + (a-1)Q^\Gamma, \quad P, Q \geq 0, a \in [0,1].$$

EW's that do not admit this form are called indecomposable.

Remark 7.19. Note that DEWs detect only states which have non-positive partial transposition (NPT). For the detection of entangled states which are PPT, one has to use indecomposable EWs.

Remark 7.20. The characterization of EWs (in particular, with respect of the notion of optimality) is still incomplete, even in the case of DEWs.

Here we give some partial results [**4, 38**]:
A sufficient criterion for the optimality of EWs is

Proposition 7.2. Let $\mathcal{H}$ be a Hilbert space of dimension $m \otimes n$ and let us define

$$P_W = \{|e, f\rangle \in \mathcal{H} : \langle e, f|W|e, f\rangle = 0\}.$$

If the set of product vectors P_W spans $\mathcal{H}$, then W is optimal.

Proposition 7.3. Let W be a DEW acting on a $2 \otimes n$ Hilbert space $\mathcal{H}$; the following are equivalent [**4**]:

(1) The set of product vectors $\{|e, f\rangle \in \mathcal{H} : \langle e, f|W|e, f\rangle = 0\}$ spans $\mathcal{H}$,
(2) W is optimal, and
(3) $W = Q^\Gamma$, where Q is a positive operator supported on a completely entangled subspace CES (i.e., $|\psi\rangle = Q|\phi\rangle$ is never a product vector).

The proof of the proposition can be found at [**4**].

Remark 7.21. However, already in the case $\mathcal{H} = \mathbb{C}^3 \otimes \mathbb{C}^3$ there exist DEWs for which $(iii) \not\Rightarrow (i)$. Consequently, the transparent characterization of DEWs in the qubit-qunit case of Proposition 7.3 does not hold in general.

To this end, [**4**] constructs the appropriate examples with the use of unextendible product basis (UPB) from $\mathbb{C}^3$, called *pyramid* [**37**].

Chapter 8
Quantum Cryptography

In this chapter, we give a brief introduction to cryptography and analyze the advantages of quantum cryptography. We explain some basic cryptographic protocols (BB84, Ekert, six state...) and analyze the possible eavesdropping strategies (differentiating on individual, collective and general attacks, depending on how much power is given to the eavesdropper), the PNS attack... We discuss which key rate can be extracted given a protocol and how authentication is performed. Finally, we focus on real system implementations [**36, 48, 14**].

8.1. Introduction, classical cryptography

The need to establish secrecy for the secure transmission of information dates back to 500 B.C. The Spartans already devised a method, the Skytale (a wooden rod around which a strap of parchment is wrapped; the diameter of the rod is secret): The message is written on the strap, each letter on a new twist, rendering it unreadable after unwrapping it. For example, the word $QUANTUM$ could be transformed to $QTUUAMN$. The receiver owns a Skytale of the same diameter allowing him to decipher the scrambled message. This cipher uses the *transposition* principle.

Another principle is the *substitution* principle, first found in Caesar's cipher, circa 50 B.C. In this case, we substitute each letter with another one with a fixed offset, i.e., $A \to D, B \to E, \dots, Y \to B, Z \to C$. In this case, the word $QUANTUM$ would be encoded into $TXDQWXP$.

Both archaic cryptosystems are poor. In every language, some letters appear more frequently than others, so a frequency analysis could easily break the substitution cipher, whereas the transposition principle is also easy to break.

8.1.1. The Vernam Cipher. The Vernam cipher [**68**] was developed in 1926. Its main idea is the addition of a *random secret key* to the message: Each letter of the plaintext is substituted by a number so that the message is a string of numbers m_i. For each i, a random number k_i is chosen and we obtain the ciphertext $c_i = m_i \oplus k_i$, where the sum is the bitwise XOR operation. For decryption, the receiver adds the same key and recovers the message $m_i = c_i \oplus k_i$.

One should note that the message and the key should be the same length, which is a drawback if large messages are to be sent, since large amounts of random numbers have to be previously distributed between the two parties.

Another important observation is that each key k_i must be used only once (this is why the Vernam cipher is also referred to as *one-time pad*). The reason for this is that if $c_i = m_i \oplus k_i$ and $c'_i = m'_i \oplus k_i$, then $c_i \oplus c'_i = m_i \oplus m'_i$, thus the ciphertext would reveal information about the original message.

The Vernam cipher has been proved to be perfectly secure [**59**].

8.1.2. RSA. This is an example of an *asymmetric* cryptosystem. The examples viewed so far were symmetric. In 1976, Diffie and Hellman [**19**] proposed the use of one-way functions for constructing an asymmetric cryptosystem. Its principle is the use of a public key, announced to everybody and a corresponding private key, kept secret. The ciphertext is computed via a *trapdoor function*, i.e., a function easy to evaluate, but hard to invert, unless having some additional information, which would be the private key.

In 1978, Rivest, Shamir and Adleman utilized this suggestion to exploit the hardness of factoring large numbers [**55**]. Their cryptosystem, now known as RSA, is still widely used in everyday life, albeit no rigorous proof of its security has been given. Furthermore, it is severely endangered by the advent of the quantum computer, as discussed in Section 6.4.4.

Explicitly, RSA works as follows: We choose two (large) prime numbers, not of very similar bit length, p_1 and p_2, and compute $N = p_1 p_2$. The Euler function of N is $\phi(N) = (p_1 - 1)(p_2 - 1)$. We choose e such that $1 < e < \phi(N)$ and $\gcd(e, N) = 1$, which is an easy task, using Euclid's algorithm. Bézout identity allows us to easily find a d such that $ed = 1 \mod \phi(N)$. The public key will be $\{N, e\}$ and the private key will be d. A message is encoded into $C = M^e \mod N$. As $a = a^{de} \mod N\ \forall a$ (this is Fermat's Theorem), the decoding can be done easily by calculating $C^d \mod N = M$.

In practice, this protocol is used for exchanging keys and then the information is encrypted using faster algorithms such as AES (Advanced Encryption Standard) which use such keys.

As a simple example, if we take $p_1 = 11, p_2 = 13, N = 143, \phi(N) = 120$. Then a pair of keys could be $\{e, N\} = \{23, 143\}$ and $\{d\} = \{47\}$.

8.2. Quantum cryptographic protocols

Quantum cryptography involves several areas of research: for example, do one-way functions exist for a quantum computer? Can RSA be generalized against adversaries who possess a quantum computer? These are some of the questions being investigated in this field. However, perhaps the most successful application of quantum cryptography is the problem of distributing a secret key through a quantum channel.

In the context of quantum key distribution (QKD), quantum states are used as information carriers. However, the term may be somewhat misleading, since it does not refer to quantum cryptosystems, but to establish a random secret key using quantum signals. As described in the previous section, the Vernam cipher is provably secure and it provides a candidate for a perfect cryptosystem, if the key distribution problem can be solved. There are two main approaches to QKD: The "prepare and measure" scheme (classical bits are encoded in a set of non-orthogonal quantum states) and the "entanglement-based" scheme (if the entanglement is maximal, simultaneous measurements will lead to perfectly correlated secret bits).

8.2.1. BB84. Named after Bennett and Brassard [**9**] in 1984, the main idea is to employ two pairs of orthogonal quantum states, where the classical bit values 0 and 1 are encoded into one pair at a time. The quantum states of one pair are non-orthogonal to the states of the other pair.

More explicitly, we take the eigenstates of the Pauli operator σ_z, which we will call $|0_z\rangle$ and $|1_z\rangle$, and the eigenstates of the Pauli operator σ_x, which we will name $|0_x\rangle$ and $|1_x\rangle$. These states share the property that $|\langle i_x|j_z\rangle| = 1/\sqrt{2}$, $i, j \in \{0, 1\}$. They are related by the Hadamard transform H: $|0_x\rangle = H|0_z\rangle$ and $|1_x\rangle = H|1_z\rangle$.

Alice and Bob are connected via a quantum channel which is totally insecure. This means that it can be assumed to be under full control of the eavesdropper, Eve. In addition, they have a public classical channel, which is authenticated (i.e., the identity of Alice and Bob is guaranteed by means of some previously shared secrecy) thus preventing Eve from sending messages impersonating Alice or Bob.

The BB84 works as follows:

1. *Preparation.* Alice prepares $2n$ qubits, each one picked at random from the set of four states $\{|0_z\rangle, |0_x\rangle, |1_z\rangle, |1_x\rangle\}$.
2. *Measurement.* For each qubit that Bob receives, he chooses at random one of the two bases z or x and measures the qubit with respect to that basis.
3. *Sifting.* Alice uses the classical channel to tell Bob in which basis she encoded each qubit. The bits where Bob used the same basis as Alice (which we expect to be n approximately) form the *sifted key.*
4. *Parameter estimation.* Alice and Bob use a subset of the sifted key to estimate the error rate. They do so by publicly announcing the bit values of the subset. If they differ in too many[1] cases, the protocol is aborted (the eavesdropper has been detected).
5. *Establishment of secret key.* Alice and Bob obtain a joint secret key from the remaining bits by performing classical error correction and privacy amplification.

Example 8.1. We present an example of the BB84 protocol. Y, N and R stand for Yes, No, Random, respectively. + is $\{|0_z\rangle, |1_z\rangle\}$ basis encoding and × is $\{|0_x\rangle, |1_x\rangle\}$ basis encoding.

Alice's string	1	0	0	1	1	1	1	0	1	0	0	0	1	1	1	1
Alice's basis	+	+	+	×	×	+	×	+	×	×	+	+	×	+	×	+
Bob's basis	+	×	+	×	×	×	+	+	×	+	+	+	+	+	×	+
Bob's string	1	R	0	1	1	R	R	0	1	R	0	0	R	1	1	1
Same basis?	Y	N	Y	Y	Y	N	N	Y	Y	N	Y	Y	N	Y	Y	Y
Bits to keep	1		0	1	1			0	1		0	0		1	1	1
Test	Y		Y	N	N			Y	Y		N	Y		N	N	Y
Key				1	1						0			1	1	

Error correction and privacy amplification are purely classical sub-protocols. Let us sketch the idea: Error correction is used to eliminate errors in the sifted key (which might be due to faulty devices, noise and/or Eve's tampering with the quantum signals). A simple error correction protocol could be as follows: Alice chooses two bits from the sifted key and tells Bob its XOR value. Bob tells Alice if the value coincides, in which case they keep the first bit and discard the second; otherwise, both bits are discarded.

[1] In further sections we will elaborate the meaning of "too many".

After error correction, once Alice and Bob share an identical bit string, the goal is to decrease Eve's knowledge about these bits. This is achieved by means of privacy amplification, a procedure which could work as follows: Alice and Bob agree on pairs of the bits of the error-corrected key and replace them with its XOR value. By doing so, the length of the key is halved, but Eve's information about this shorter key is less; even if she knew the values of the single bits with high probability.

Eavesdropping Strategies and Disturbance Versus Information Gain

So far we have given a description of the BB84 protocol, but we haven't analyzed its security. The most simple attack Eve could perform is an *intercept and resend* attack of the $2n$ qubits that travel from Alice to Bob. Since Theorem 4.1 forbids the perfect copy of quantum states, an obvious strategy is to measure them. However, Eve does not know which basis they were originally prepared in, since Alice announces this information once Bob has received all signals. All Eve can do is guess, i.e., for about n qubits she will happen to choose the same basis as Alice and get perfectly correlated results. In the other half, her results will be completely random and uncorrelated.

Then, Eve has to resend these bits to Bob, but she does not know which basis Alice chose, so Eve prepares each qubit in the same basis used for the measurement. This means that only about n of the newly created qubits will match Alice's bases. Bob receives Eve's qubits and measures them. Afterward, Bob and Alice apply the sifting phase 3.

Alice and Bob's bases will be the same and Eve's basis will be different in approximately $n/2$ cases. In such cases, Bob's result will be random, which means that the sifted key will be wrong for about $n/4$ bits. If, in the parameter estimation stage 4., Alice and Bob obtain such a high error rate (25%) the protocol is aborted.

Remark 8.1. This simple example shows that the intercept & resend strategy induces Eve to introduce errors when she learns parts of the key.

In general, there is a trade-off between disturbance and information gain. We shall illustrate this fact in this more (yet not most) general strategy:

Now Eve possesses an ancillary system $|E\rangle$ and a unitary interaction. Let U denote the unitary operation employed by Eve.

- If Eve does not disturb Alice's qubits, the action on two non-orthogonal states among $\{|0_z\rangle, |0_x\rangle, |1_z\rangle, |1_x\rangle\}$ is given by
$$U|0_z\rangle|E\rangle = |0_z\rangle|E_{0_z}\rangle,$$
$$U|1_x\rangle|E\rangle = |1_x\rangle|E_{1_x}\rangle,$$
where the notation is self-explanatory, and we can consider this case without loss of generality. Since unitary operations preserve the scalar product, the scalar product of the right-hand sides and left-hand sides are
$$\langle 0_z|1_x\rangle\langle E|E\rangle = \langle 0_z|1_x\rangle\langle E_{0_z}|E_{1_x}\rangle$$
This implies that $|E_{0_z}\rangle$ and $|E_{1_x}\rangle$ must be identical; thus, Eve cannot gain any information when measuring her ancilla.
- Let us suppose that Eve's attack does disturb the qubits that Alice sends. In this case, we use the notation
$$U|0_z\rangle|E\rangle = |0'_z\rangle|E_{0_z}\rangle,$$

$$U|1_x\rangle|E\rangle = |1'_x\rangle|E_{1_x}\rangle,$$

which implies that

$$\langle 0_z|1_x\rangle\langle E|E\rangle = \langle 0'_z|1'_x\rangle\langle E_{0_z}|E_{1_x}\rangle.$$

In order to allow Eve to obtain information about the states sent by Alice, she needs to make the states $|E_{0_z}\rangle$ and $|E_{1_x}\rangle$ distinguishable. This means that their scalar product must decrease, so $\langle 0'_z|1'_x\rangle$ must increase.

Remark 8.2. This shows that the more information Eve wants to obtain, the more disturbance she introduces to the signal.

8.2.2. Six-state. The six-state protocol [**12**] is a variant of the BB84 protocol described in the previous section. The enhancement is to introduce the eigenvalues of σ_y into the encoding. Thus, we encode in three mutually unbiased bases of the two-dimensional Hilbert space: $\{|0_\alpha\rangle, |1_\alpha\rangle\}_{\alpha\in\{x,y,z\}}$ with $|\langle i_\alpha|j_\beta\rangle| = 1/\sqrt{2}$ if $\alpha \neq \beta$ and $i, j \in \{0, 1\}$.

Intuitively speaking, this protocol is better since the six states span the full Bloch sphere, rather than only a great circle, which would be the case of BB84. Indeed, Alice can choose one out of three encodings and in the sifting procedure, approximately 2/3 of the raw key bits get discarded. Thus, a higher secret key rate can be extracted, since the eavesdropper has less prior information.

8.2.3. Ekert. The Ekert protocol (1991) [**22**] uses entanglement to create a secret key for Alice and Bob. The idea is to distribute maximally entangled singlet states

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

The basic feature to be exploited is the fact that a measurement o both qubits in any basis yields correlated results. The problem is now the distribution of the qubits, such that one can be sure that an eavesdropper can get no (or very limited) information about the final key. This is accomplished by checking a Bell inequality; namely, the CHSH inequality described in 4.1.1.

The protocol is stated as follows:

1. *Entanglement distribution.* Alice and Bob distribute a number of singlet states $|\psi^-\rangle$ among them. We shall suppose that the first subsystem belongs to Alice and the second belongs to Bob.
2. *Measurements.* For each singlet, Alice and Bob measure an observable, which is randomly chosen from the sets $\{A_i\}_i$ and $\{B_i\}_i$, respectively. These observables are spin components, lying on the $x - z$ plane of the Bloch sphere, given by
$$A_i = \cos\phi_i^A\sigma_z + \sin\phi_i^A\sigma_x, \quad B_i = \cos\phi_i^B\sigma_z + \sin\phi_i^B\sigma_x,$$
where the angles are $\phi_1^A = 0, \phi_2^A = \pi/2, \phi_3^A = \pi/4$ for Alice and $\phi_1^B = 0, \phi_2^B = -\pi/4, \phi_3^B = \pi/4$ for Bob. In Fig. 8.1 we have illustrated a graphical representation of the measurement directions.
3. *Announcement of bases.* Alice and Bob announce the directions they chose for each measurement. In the cases where their directions match; namely, (A_1, B_1) and (A_3, B_3), their results are completely anticorrelated. The sifting key is formed by inverting all bits from one party.
4. *Bell inequality test.* The CHSH inequality is tested, using the results obtained when Alice and Bob measured in the directions $(A_1, B_3), (A_1, B_2), (A_2, B_3)$ and (A_2, B_2).
5. *Establishment of secret key.* Alice and Bob obtain a joint secret key from the sifted key by means of error correction and privacy amplification, as described in the BB84 protocol.

Step 4. is what guarantees if Alice and Bob share a maximally entangled state (thus, it cannot be entangled between any under possession of Eve; i.e., Eve has no information on the key).

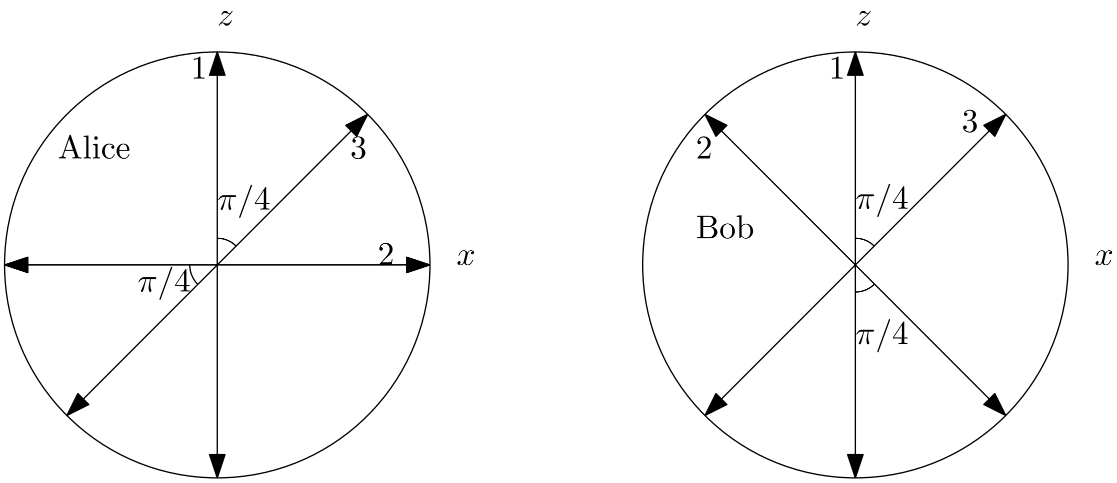

FIG. 8.1. Ekert protocol. Alice's and Bob's measurement directions on the Bloch sphere.

Remark 8.3. In this case, the CHSH inequality would be as follows[2]:

If we have four classical random variables A_1, A_2, B_2, B_3 taking values on $\{-1, 1\}$, it is trivial to see that $A_1(B_3 + B_2) + A_2(B_3 - B_2) = \pm 2$. Taking the average over N assignments to these variables, one finds that

$$|\langle A_1(B_3 + B_2) + A_2(B_3 - B_2)\rangle| \leq 2.$$

In the form of a CHSH inequality, this would be

$$S := |\langle A_1B_3\rangle + \langle A_1B_2\rangle + \langle A_2B_3\rangle - \langle A_2B_2\rangle| \leq 2,$$

where $\langle A_iB_j\rangle = \frac{1}{N}\sum_{\nu\in I} A_i^{(\nu)}B_j^{(\nu)}$, $\#I = N$, and $A_i^{(\nu)}$ is assignment number ν of variable A_i.

If the variables are quantum mechanical observables their expectation number would now be

$$\langle A_iB_j\rangle = Tr(A_i \otimes B_j\rho).$$

In the case of Ekert protocol, $\rho = |\psi^-\rangle\langle\psi^-|$. Using directions described in step 2., the value of S is $S = 2\sqrt{2}$, a violation of the CHSH inequality.

So, whenever Alice and Bob measure $S = 2\sqrt{2}$, they can be sure to share a maximally entangled state between them. If the entanglement is maximal, it cannot be entangled with anything else; in particular, anything under Eve's control.

On the other hand, if no violation of the CHSH is found, Alice and Bob's measurement are compatible with a separable state, rendering impossible to create a secret key. In this case, the protocol should be aborted.

[2] We already derived the inequality in Section 4.1.1, but we present a much clearer approach here which needs not involve a Hidden Variable Model.

Remark 8.4. We want to emphasize that the protocols mentioned for QKD have the problem of authentication. Indeed, Bob must know that it is really Alice who is talking (for example, at the sifting procedure) or the whole security of the protocol could be compromised. We analyze the different kinds of strategies a spy could perform to break a QKD cryptosystem in the following section. The authentication procedure is still an area of active research today [**36**].

8.3. Eavesdropping strategies

In a formal manner, the set of all eavesdropping strategies can be classified into three classes, which we will call *individual*, *collective* and *coherent* attacks. The criteria for distinguishing them is the power we give to the eavesdropper.

The most general way to describe how information can be extracted from a quantum system in a state ρ_A is the following [**14**]: Attach an ancilla system in a predefined known state $|0\rangle\langle 0|_E$; then perform a -sophisticated- unitary operation U on the composite system $\rho_A \otimes |0\rangle\langle 0|_E$. Then, do an -also sophisticated- measurement on the ancilla system, which is

$$\rho_E = Tr_A(U^\dagger(\rho_A \otimes |0\rangle\langle 0|_E)U).$$

The measurement to be performed is given by a POVM $\mathcal{M}$ described by the projections on the eigenspaces $\{M_j\}_j$, and which yields outcome j with probability $Tr(M_j\rho)$ when measuring a state ρ.

Notation 8.1. We shall denote the classical probability distribution induced by the measurement by $\mathbf{P}^{\rho}_{\mathcal{M}}(j) = Tr(M_j\rho)$.

If we want to indicate the number of subsystems on which the measurement is performed, we will write $\mathcal{M}^1$ or $\mathcal{M}^n$.

Let us consider the case where Alice sends n quantum systems $\rho_A^1, \dots, \rho_A^n$ to Bob.

- Individual attacks are the simplest ones, and correspond to an eavesdropper with little power.
  More precisely, Eve will attach an ancilla system $|0\rangle\langle 0|_E$ to each state ρ_A^i and applies the same unitary operation, U, performing the same measurement on her part of all the composite systems individually and in the same way.
  The probability distribution that Eve obtains for each class of attacks is given by
  $$\mathbf{P}^{\rho_E^1}_{\mathcal{M}^1} \cdots \mathbf{P}^{\rho_E^n}_{\mathcal{M}^1}, \quad \text{where} \quad \rho_E^i = Tr_A(U^\dagger(\rho_A^i \otimes |0\rangle\langle 0|_E)U).$$
- Collective attacks are a bit more general, as they allow the eavesdropper to measure all ancilla systems collectively.
  In this case, the probability distribution obtained by Eve will be
  $$\mathbf{P}^{\rho_E^1 \otimes \cdots \otimes \rho_E^n}_{\mathcal{M}^n}, \quad \text{where} \quad \rho_E^i = Tr_A(U^\dagger(\rho_A^i \otimes |0\rangle\langle 0|_E)U).$$
- Coherent attacks are potentially the most powerful, since they correspond with an eavesdropper with unlimited technological power and resources, only limited by laws of Nature.
  In a coherent attack, Eve attaches one large ancilla system to the state $\rho_A^1 \otimes \cdots \otimes \rho_A^n$ and then performs a *global* unitary transformation U_g and measurement $\mathcal{M}^n$.
  The probability distribution obtained by Eve will contain the most information in this case:
  $$\mathbf{P}^{\rho_E}_{\mathcal{M}^n}, \quad \text{where} \quad \rho_E = Tr_A(U_g^\dagger((\rho_A^1 \otimes \cdots \otimes \rho_A^n) \otimes |0\rangle\langle 0|_E)U_g).$$

Remark 8.5. Coherent attacks are not only the most general ones; they also are the most difficult ones to analyze, due to the high dimension of the global Hilbert space, which grows exponentially. Coherent eavesdropping has been studied for the BB84 [**14**] and for the six-state protocol [**14**]. As a quite surprising result, the authors found that coherent eavesdropping does not increase Eve's Shannon information; however, it does slightly increase the probability to guess the key.

8.3.1. PNS attack. The different kinds of attacks considered so far involve idealized QKD protocols and eavesdropping.

Nevertheless, realistic experimental implementations may offer Eve new and more powerful paths. When looking at a common experimental implementation for QKD, a dangerous strategy becomes obvious:

Qubit systems can be conveniently represented by photons, typically using their polarization as a degree of freedom. Ideally, each qubit is encoded by exactly one photon. However, single-photon sources do not exist. Instead, one often uses weak coherent pulses,

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle,$$

where $\alpha \in \mathbb{C}$ and $|n\rangle$ is a Fock state, as described in Section 5.2.2.

If the phase of α, $\arg\alpha$ is unknown or it is randomized, one arrives at the following mixture of Fock states [**14**]:

$$\rho = \int \frac{1}{2\pi}|\alpha\rangle\langle\alpha| d\arg\alpha = \sum_{n=0}^{\infty}\mathbf{P}(n)|n\rangle\langle n|,$$

where the probability distribution of photons obeys Poissonian statistics, i.e.,

$$\mathbf{P}(n) = e^{-|\alpha|^2}\frac{|\alpha|^{2n}}{n!}.$$

The mean photon number is $\overline{n} = |\alpha|^2$. Thus, a *weak* laser pulse contains no photons in most cases, and more than a photon with a small probability. A typical value used for QKD [**14**] has a mean photon number $\overline{n} = 0,1$. This corresponds to $\mathbf{P}(0) \simeq 90,5\%$ probability of no photons; $\mathbf{P}(1) \simeq 9\%$ of one photon; thus, the probability for more than a photon would be $\simeq 0,5\%$.

The photon number splitting (PNS) attack is as follows: For each weak coherent pulse sent from Alice to Bob via a lossy optical fiber, Eve measures its photon number (via a non-demolition measurement). If it is more than one, she splits off one photon, stores it and sends the remaining photon(s) to Bob via a lossless fiber. If the photon number is one, she blocks the event with certain probability (according to the photon statistics Bob expects to receive). Vacuum events are just forwarded to Bob.

For all multi-photon events, Eve can get full information on the corresponding key bit if she waits for Alice to announce the bases, measuring correctly each corresponding photon. Thus, the protocol loses its unconditional security.

Several strategies have been proposed to counter this attack: For example, the use of *decoy states* [**14, 36**], which consists of the introduction of a second source in Alice's side sending weak coherent pulses with different photon number distribution; namely, the decoy source has a much higher mean $\overline{n}_{dec}$, yet it is not different from the signal source in any other parameter. After sending all pulses, Alice announces which were prepared by each source. Eve can be detected in case she launched a PNS attack by Bob, who will find an abnormally higher loss for the photons with lower mean photon number $\overline{n}$.

Remark 8.6. Other kinds of attacks in a realistic environment are possible [**36**] besides the PNS attack. We shall only sketch its main idea here.

Example 8.2. For example, the *Trojan Horse* attack (or light injection attack) [**36**]: Eve aims not to interact with the photons in transit between Alice and Bob, but to probe the device in Alice's and Bob's side by sending some light into them and collecting the reflected signal.

Eve can use the information of the reflected signal to detect which basis Alice used for the preparation of the photon, via a phase modulation due to the different ways the reflected and the reference pulse can go through. If Eve is able to perform this before Alice's photon reaches Bob, she can perform a simple intercept & resend (I&R) attack and gain full information on the secret bit string, since Eve can always measure in the correct basis.

A countermeasure could be the use of isolators in Alice's lab (when the communication is from Alice to Bob) and monitor the intensity of incoming light.

Example 8.3. The *faked states* attack is a kind of I&R strategy in which Eve does not try to intercept the recreated state. Instead, Eve manages to send a signal to Bob which can only be detected in a way totally controlled by Eve. The *time-shift* strategy is an alternative version of this [**36**].

8.4. Bounds

Definition 8.1. Let A, B, E be random variables with joint probability distribution P_{ABE}. The conditional mutual information between A and B, given E, is defined as

$$I(A;B|E) = \sum_{e\in\mathcal{E}} P_E(e)[H(A|e) + H(B|e) - H(A,B|e)],$$

where the conditional Shannon entropy is defined as

$$H(X|e) = -\sum_{x\in\mathcal{X}} P_{X|E}(x,e)\log P_{X|E}(x,e),$$

and we consider that E takes values in $\mathcal{E}$ and X in $\mathcal{X}$.

Remark 8.7. The conditional mutual information $I(A;B|E)$ quantifies the amount of information revealed about A when learning B, given the knowledge of E.

When one minimizes $I(A;B|E)$ over all the possible processing of the variable E obtains

Definition 8.2. The intrinsic information between A and B, given E, is defined as

$$I(A;B\downarrow E) = \inf_{E\longrightarrow\tilde{E}} I(A;B|\tilde{E}),$$

where the infimum is taken over all channels $P_{\tilde{E}|E}$ that can be used to process E.

The following definitions quantify how many secret bits can be extracted from a given probability distribution and how many are needed to create it:

Definition 8.3. The secret key rate $S(A:B||E)$ is defined as the maximal amount of secret bits that can be extracted asymptotically from P_{ABE}.

Definition 8.4. The information of formation $I_{form}(A;B|E)$ is defined as the minimal number of secret bits that are needed asymptotically to create P_{ABE}.

Proposition 8.1. The following bounds of the secret key rate can be derived:

$$S(A;B||E) \leq I(A;B \downarrow E) \leq I_{form}(A;B|E).$$

The proof of this proposition can be found at [**14**].

8.5. Certified random number generation

In this last section, we present a particular experiment [**52**] with cryptographic applications: the problem of certifying a sequence is genuinely random.

Randomness is a fundamental feature of nature and a valuable resource for multiple applications (numerical simulation of physical and/or biological systems, cryptography...). However, the characterization of true randomness is elusive. There exist a series of statistical tests [**57**] which are designed to look for the existence of patterns that may appear in non purely random sequences. Yet, not all patterns can be covered by a finite number of such tests.

At a more fundamental level, in the classical world, any system admits in principle a deterministic description and apparent randomness is due to our lack of knowledge. On the other hand, quantum theory is intrinsically random.

Let us consider a RNG (random number generator) as a black box. Let us also assume the possibility that an adversary has tampered with the RNG introducing patterns undetected by any statistical test. The patterns we consider also include the case that all the numbers were generated in advance by the adversary and copied into a memory inside the black box.

Is there any means to guarantee the existence of *private* randomness?

The laws of quantum mechanics enable procedures to achieve that. In the experiment proposed at [**52**], the violation of Bell's inequalities (like CHSH described in Section 4.1.1) is used to certify non-local correlations and to generate private randomness.

Let us begin by describing the experiment: Fig. 8.2 shows the experimental setup. The process begins with each atom emitting a single photon (to the left). Each photon is entangled with the corresponding atom. The aim is to entangle the two $^{171}\text{Yb}^{+}$ atoms. In order to achieve this, the procedure used is known as *entanglement swapping*: roughly speaking, the two photons are entangled and then this entangled is transferred to the atoms. Hence, the photons are collected with respective optical fibers and they interfere in the beamsplitter (BS). Finally, each photon is measured with a photo-multiplier tube (PMT). At this stage, the two qubits are entangled. Random binary inputs x, y are given to microwave oscillators, which coherently rotate each qubit accordingly (in plain terms, x, y determine in which basis we will measure each qubit). To ensure no influence of one measurement to another, during the measurement process, the two boxes were not allowed to communicate. The measurement outputs binary values a and b. More precisely, atomic fluorescence techniques were used (with detection error probability $< 3\%$).

The Bell inequality was quantified through the CHSH correlation function [**16, 52**]

$$I = \sum_{x,y} (-1)^{xy} (\mathbf{P}(a = b|xy) - \mathbf{P}(a \neq b|xy)),$$

with an estimator for I for n samples:

$$\hat{I} = \frac{1}{n} \sum_{x,y} (-1)^{xy} (N(a = b|xy) - N(a \neq b|xy))/\mathbf{P}(xy),$$

where N counts the corresponding number of events.

The total number of samples was $n = 3.016$, collected over the period of one month (this is due to the low probability of success of entanglement generation, $\mathbf{P}(ent) \sim 2 \cdot 10^{-8}$).

The experimental results are shown in the following table:

(x, y)	(ϕ_x, ϕ_y)	$N_{(0,0;x,y)}$	$N_{(0,1;x,y)}$	$N_{(1,0;x,y)}$	$N_{(1,1;x,y)}$	Total	$\widehat{\mathbf{P}}(a = b\|xy)$
$0, 0$	$0, \pi/4$	293	94	70	295	752	$0, 782$
$0, 1$	$0, 3\pi/4$	298	70	74	309	751	$0, 808$
$1, 0$	$\pi/2, \pi/4$	283	69	64	291	707	$0, 812$
$1, 1$	$\pi/2, 3\pi/4$	68	340	309	89	806	$0, 195$

In order to minimize the number of runs to obtain a meaningful bound, the input was distributed as $\mathbf{P}_{X,Y} \sim$Uniform$(1/4)$. Also, for the collected data, $\hat{I} = 2, 414 \pm 0, 058$. Hence, the probability that these results could be explained by means of a local theory ($I \leq 2$) is $\mathbf{P}(\hat{I} \geq 2, 414) \leq 0, 00077$.

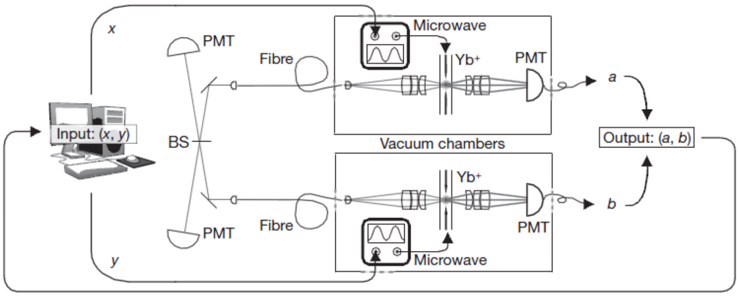

FIG. 8.2. Experimental realization of private random number generator using two ^{171}Yb^{+} qubits trapped in independent vacuum chambers.

The CHSH violation observed implies that > 42 new random bits are generated with a 99% confidence level [**52**].

Chapter 9
Quantum Coding

In this chapter, we treat the subject of quantum coding and quantum error correction. We treat the different kinds of decoherence a quantum channel can undergo. We give some examples of quantum codes (three-qubit bit-flip/phase-flip codes, nine-qubit Shor code) and introduce CSS codes. We analyze how adding extra qubits does in fact compensate the decoherence they introduce in the system. Finally we treat the case of degenerate codes [**8, 44, 48, 50**].

9.1. Introduction

In practice, any quantum system is open; namely, it is never perfectly isolated from the environment. The core of the problem is the superposition principle, according to which, any superposition of quantum states is an acceptable quantum state. We use the word decoherence in order to refer to any quantum-noise process due t coupling a system with the environment.

In quantum information processing, decoherence is the major threat to the actual implementation of any quantum computation or communication protocol.

9.2. Decoherence

We shall begin this section by studying quantum-noise (decoherence) processes that can act on a single qubit, before giving the general formulation of the problem in Section 9.2.1. We shall consider the environment as a single qubit and the system-environment ($\Sigma + E$) interaction shall be modeled as a CNOT gate (Fig. 9.1). Let us assume that initially the system Σ is in a pure state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with its corresponding density matrix representation in the basis $\{|0\rangle, |1\rangle\}$ and that E is in the initial state $|0\rangle$:

$$\rho = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}$$

As seen in Remark 3.9, the diagonal terms are known as populations and represent the probabilities to obtain the outcomes 0 or 1 from a measurement along the z-axis, respectively. The off-diagonal terms, known as coherences, appear when $|\psi\rangle$ is a superposition of the states $|0\rangle$ and $|1\rangle$ and shall be completely destroyed by the decoherence process depicted in Fig. 9.1.

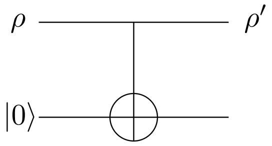

FIG. 9.1. Quantum circuit modeling decoherence process.

Indeed, the initial global $\Sigma + E$ state

$$|\Psi\rangle = |\psi\rangle \otimes |0\rangle = (\alpha|0\rangle + \beta|1\rangle)|0\rangle$$

evolves into he final state

$$|\Psi'\rangle = \alpha|00\rangle + \beta|11\rangle. \tag{9.1}$$

Remark 9.1. The CNOT interaction has entangled the qubit with the environment, as the state $|\Psi'\rangle$ is non-separable. By tracing over the environment, we obtain the final density matrix ρ' of the system Σ:

$$\rho' = Tr_E|\Psi\rangle\langle\Psi| = \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix}.$$

Remark 9.2. This decoherence process allows for the following interpretation: From (9.1) it is obvious that the environment has learned, through the CNOT interaction, what is the state of Σ (if Σ is in the state $|0\rangle$, E remains in the state $|0\rangle$; if Σ is in the state $|1\rangle$, E flips to the state $|1\rangle$). Therefore, we can think of the CNOT gate as a measurement performed by E onto Σ.
Now the information of the relative phases of the coefficients α, β is encoded into the $\Sigma + E$ quantum correlations. As we do not keep record of the state of the environment E, this information is eventually lost for us.
In short, information leaks from the system into the external world.

Example 9.1. *Quantum circuits simulating noise channels.*

In a more general picture, let us consider a system Σ described by a density matrix ρ and an environment E with two ancillary qubits, in the pure state

$$|\psi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle,$$

with $\alpha, \beta, \gamma, \delta \in \mathbb{C}$ satisfying the normalization condition $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$. The initial density matrix $E + \Sigma$ will have the following block representation:

$$\rho_{(E+\Sigma)} \equiv \widetilde{\rho} = |\psi\rangle\langle\psi| \otimes \rho = \begin{pmatrix} |\alpha|^2\rho & \diamond & \diamond & \diamond \\ \diamond & |\beta|^2\rho & \diamond & \diamond \\ \diamond & \diamond & |\gamma|^2\rho & \diamond \\ \diamond & \diamond & \diamond & |\delta|^2\rho \end{pmatrix},$$

where $\diamond$ denotes blocks not needed in subsequent calculations.

Consider the quantum circuit which implements the unitary transform with the following block representation:

$$U = \begin{pmatrix} \sigma_x & 0 & 0 & 0 \\ 0 & \sigma_y & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & \mathbb{I}_2 \end{pmatrix}$$

Thus, the final state $E + \Sigma$ is described by $\widetilde{\rho}' = U\widetilde{\rho}U^\dagger$, which is, in general, entangled with the environment.

To obtain ρ' we trace out the environmental qubits

$$\rho' = Tr_E\widetilde{\rho}' = |\alpha|^2\,\sigma_x\rho\sigma_x^\dagger + |\beta|^2\,\sigma_y\rho\sigma_y^\dagger + |\gamma|^2\,\sigma_z\rho\sigma_z^\dagger + |\delta|^2\,\rho,$$

which can be expressed through the Kraus operators introduced in Remark 7.7 $E_1 = |\alpha|\,\sigma_x$, $E_2 = |\beta|\,\sigma_y$, $E_3 = |\gamma|\,\sigma_z$ and $E_0 = |\delta|\,\mathbb{I}_2$ leading to the Kraus representation

$$\rho' = \sum_{i=0}^{3} E_i\rho E_i^\dagger. \tag{9.2}$$

Remark 9.3. The transformation induced by the Kraus operators in (9.2) has a geometric representation on the Bloch sphere. Let us associate the Bloch vectors $\boldsymbol{r}, \boldsymbol{r'}$ with the density matrices ρ, ρ' respectively.

Using the Bloch sphere representation (3.3), direct computation shows that

$$2\sigma_x\rho\sigma_x^\dagger = \begin{pmatrix} 1-z & x+iy \\ x-iy & 1+z \end{pmatrix}, \quad 2\sigma_y\rho\sigma_y^\dagger = \begin{pmatrix} 1-z & -(x+iy) \\ -(x-iy) & 1+z \end{pmatrix}$$

and $2\sigma_z\rho\sigma_z^\dagger = \begin{pmatrix} 1+z & -(x-iy) \\ -(x+iy) & 1-z \end{pmatrix}$.

Using the normalization condition $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$ we obtain

$$\begin{aligned} x' &= (1 - 2(|\beta|^2 + |\gamma|^2))x, \\ y' &= (1 - 2(|\alpha|^2 + |\gamma|^2))y, \\ z' &= (1 - 2(|\alpha|^2 + |\beta|^2))z. \end{aligned}$$

This tells us that the Bloch sphere is deformed into an ellipsoid centered at the origin of the Bloch sphere and with axes directed along x, y, z.

Remark 9.4. Depending on the choice of parameters $|\alpha|, |\beta|, |\gamma|$ (observe that their relative phase does not affect the final state ρ) we can obtain various commonly investigated channels, such as the bit-flip channel already seen in Example 7.1 with $\beta = \gamma = 0$ or their analogous phase-flip channel with $\alpha = \beta = 0$ and bit-phase-flip channel with $\alpha = \gamma = 0$. The depolarizing channel introduced in Example 7.2 is obtained by setting $|\alpha|^2 = |\beta|^2 = |\gamma|^2 = p/3$, with its Bloch sphere representation $\boldsymbol{r'} = \left(1 - \frac{4}{3}p\right)\boldsymbol{r}$ resulting in a shrinking of the sphere.

Remark 9.5. The models we have seen in this section and in Example 7.4 leading to decoherence are phenomenological. They do not represent the physical mechanisms inducing decoherence any better than a resistance in an electric circuit represents the scattering process undergone by electrons. We justify decoherence by means of a simple model which allows more formal developments [**8**].

9.2.1. Quantum to classical transition. Since the dawn of quantum theory, the emergence of classical behaviour in a world governed by the laws of quantum mechanics has been a fascinating problem. The heart of the problem is the superposition, which entails consequences that appear unacceptable to classical intuition, as we have seen for example in the CHSH inequality (4.1) discussion and Example 4.1.

Example 9.2. *Schrödinger's cat.*
Schrödinger's cat paradox is a −theoretical− experiment devised by the Austrian physicist Erwin Schrödinger in 1935 to discuss the Copenhagen interpretation of quantum mechanics and the EPR article [**23**] which had highlighted the strangeness of the subject.

Inside a box we have a radioactive source, a detector, a hammer, a vial of poison and a cat. The source is a two-level atom Σ, initially in its excited state $|1\rangle$ (we shall use the model already described in Example 7.3). The atom can decay to the ground state $|0\rangle$ by emission of a photon, which triggers the detector. The clock of the detector induces the hammer to break the vial and kill the cat (system Σ', where $|\star\rangle$ shall mean the cat is alive and $|\dagger\rangle$ it is dead).

We assume that the composite $\Sigma + \Sigma'$ system is

$$|\psi_0\rangle = |1\rangle|\star\rangle.$$

After the half-life of the atom we have the state $|\psi\rangle = \frac{1}{\sqrt{2}}\left(|1\rangle|\star\rangle + |0\rangle|\dagger\rangle\right)$, a superposition of the live and dead cat states. Note that now the cat and the atom are entangled.

More precisely, the density matrix of the state $|\psi\rangle$ is

$$\rho = |\psi\rangle\langle\psi| = \frac{1}{2}\left(|1\rangle|\star\rangle\langle 1|\langle\star| + |0\rangle|\dagger\rangle\langle 0|\langle\dagger| + |1\rangle|\star\rangle\langle 0|\langle\dagger| + |0\rangle|\dagger\rangle\langle 1|\langle\star|\right),$$

which has the representation in the basis $\{|0\rangle|\star\rangle, |0\rangle|\dagger\rangle, |1\rangle|\star\rangle, |1\rangle|\dagger\rangle\}$:

$$\rho = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

This density matrix contains non-zero matrix elements along the diagonal and also off-diagonal. The latter (coherences) have no classical analogue -to the date, no cat has been found alive and dead at the same time yet-.

The paradox is solved via decoherence. Decoherence plays a key role in understanding the transition from the quantum to the classical world.

The atom-cat system is never perfectly isolated from the environment, which it is constantly measuring it. The state we must consider is (by denoting $|E_0\rangle, |E_1\rangle$ the states of the environment)

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|1\rangle|\star\rangle|E_1\rangle + |0\rangle|\dagger\rangle|E_0\rangle\right).$$

If $|E_1\rangle, |E_0\rangle$ are orthogonal states, we obtain, the final state corresponding to $\Sigma + \Sigma'$ by tracing out the environment E:

$$\rho_{\text{dec}} = Tr_E|\Psi\rangle\langle\Psi| = \frac{1}{2}\left(|1\rangle|\star\rangle\langle 1|\langle\star| + |0\rangle|\dagger\rangle\langle 0|\langle\dagger|\right),$$

which is now a diagonal matrix corresponding to a mixed state. It is now compatible with a classical description in terms of probabilities: After the half-life of the atom, the cat is dead with probability $p = 1/2$ and alive with the same probability. We discover its state upon observation.

9.3. Quantum Codes

We have already seen how the process of decoherence can lead to destruction of quantum information. To fight the effect of noise, a well-developed technique is the use of error-correcting codes. The main ingredient to achieve such aim is the use of redundancy.
The simplest way to protect a classical bit is to send three copies of it; for instance, if Alice wishes to send Bob a classical bit, she may send 000 instead of 0 or 111 instead of 1. Bob receives the three bits, maybe with some errors, and decides with majority voting which one was sent.

Remark 9.6. There is an underlying hypothesis in this reasoning, which is that the noisy channel is memoryless; namely, noise acts independently on each bit. If ϵ is the probability of error on a single bit, then the code fails with probability $\epsilon_c = 3\epsilon^2(1-\epsilon)+\epsilon^3 = 3\epsilon^2 - 2\epsilon^3$ (if two or more bits have been flipped). If $\epsilon_c < \epsilon$, this means that the code improves the possibilities of successful transmission, which happens if $\epsilon < 1/2$.

We wish to apply the same principle to quantum information; however we encounter difficulties due to the basic principles of quantum mechanics:

1. According to the no-cloning theorem it is impossible to make copies of an unknown quantum state. Therefore, we cannot just send $|\psi\rangle|\psi\rangle|\psi\rangle$ in order to protect an unknown $|\psi\rangle$.
2. In order to operate classical error correction, we observe (measure) the output from the noisy channel. In quantum mechanics, measurement disturbs (in general) the system: If we receive $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and measure its polarization along the z-axis, the state will collapse onto $|0\rangle$ or $|1\rangle$, destroying the coherent superposition of $|\psi\rangle$.
3. The only possible classical error affecting a single bit is the bit flip error $0 \rightleftarrows 1$. There is a much wider class of quantum errors, in general with no classical counterparts (e.g. the phase-flip error $\alpha|0\rangle + \beta|1\rangle \rightleftarrows \alpha|0\rangle - \beta|1\rangle$; noise may also slightly rotate a state $|\psi\rangle \to R|\psi\rangle$, with R a rotation matrix). At first sight it might appear that infinite precision is needed in order to correct this continuum of quantum errors (e.g., the rotation angle of R); however, we shall see that quantum error correction is indeed possible.

9.3.1. Three-qubit bit-flip code. We shall assume that noise acts on each qubit independently. Imagine that Alice wishes to send a qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob via a noisy quantum channel which will leave the state of the qubit unchanged with probability $1-\epsilon$ and will apply the Pauli operator σ_x with probability ϵ. Note that this channel acts like the one already described in Example 7.1.

Notation 9.1. We shall use the following notation for the logical states or codewords: $|0_L\rangle \equiv |000\rangle, |1_L\rangle \equiv |111\rangle$.

Alice will employ the encoding $|0\rangle \to |0_L\rangle, |1\rangle \to |1_L\rangle$. Thus, a generic state will be encoded as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \to \alpha|0_L\rangle + \beta|1_L\rangle.$$

A simple CNOT circuit in cascade can implement this encoding:

$$CNOT_{1,3}(CNOT_{1,2}(\alpha|0\rangle + \beta|1\rangle)|00\rangle) = CNOT_{1,3}(\alpha|00\rangle + \beta|11\rangle)|0\rangle = \alpha|0_L\rangle + \beta|1_L\rangle. \tag{9.3}$$

This state is known as the GHZ (named after Greenberger, Horne and Zeilinger) state, or cat state.

Remark 9.7. This does not violate the no cloning theorem, since $|\psi\rangle|\psi\rangle|\psi\rangle \neq \alpha|0_L\rangle + \beta|1_L\rangle$.

After passing the channel, Bob may receive the following states, with its corresponding probabilities:

$$\begin{array}{lr}
\alpha|000\rangle+\beta|111\rangle, & (1-\epsilon)^3, \\
\alpha|100\rangle+\beta|011\rangle, & \epsilon(1-\epsilon)^2, \\
\alpha|010\rangle+\beta|101\rangle, & \epsilon(1-\epsilon)^2, \\
\alpha|001\rangle+\beta|110\rangle, & \epsilon(1-\epsilon)^2, \\
\alpha|110\rangle+\beta|001\rangle, & \epsilon^2(1-\epsilon), \\
\alpha|101\rangle+\beta|010\rangle, & \epsilon^2(1-\epsilon), \\
\alpha|011\rangle+\beta|100\rangle, & \epsilon^2(1-\epsilon), \\
\alpha|111\rangle+\beta|000\rangle, & \epsilon^3.
\end{array}$$

In order to prevent a single bit-flip error, Bob could proceed as in classical error correction and measure the polarizations $\sigma_z^{(i)}$ of the three qubits and apply majority voting. However, no quantum state would remain afterward.

The problem is solved by performing collective measurements on two simultaneous qubits.

This process is illustrated in Fig. 9.2: with the aid of two extra ancillary qubits in the original state $|00\rangle$, Bob is now allowed to measure $\sigma_z^{(1)}\sigma_z^{(2)}$ and $\sigma_z^{(1)}\sigma_z^{(3)}$. Thus, Bob obtains an error syndrome x_0, x_1 from the two classical measurements.

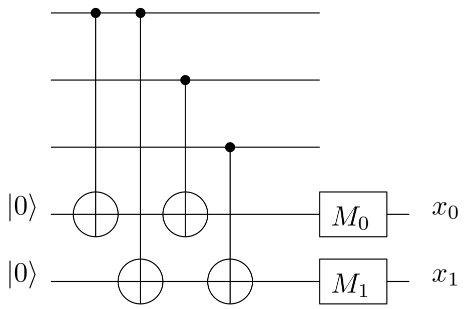

FIG. 9.2. The circuit for extraction of the error syndrome x_0, x_1 by performing measures M_0, M_1 on the ancillary qubits.

Bob's actions should be the following:

- $x_0 = x_1 = 0$. No action.
- $x_0 = 0, x_1 = 1$. Apply NOT to the third qubit.
- $x_0 = 1, x_1 = 0$. Apply NOT to the second qubit.
- $x_0 = x_1 = 1$. Apply NOT to the first qubit.

After Bob's correction, the five-qubit states and their corresponding probabilities will be

$$\begin{array}{lr}
(\alpha|000\rangle+\beta|111\rangle)|00\rangle, & (1-\epsilon)^3, \\
(\alpha|000\rangle+\beta|111\rangle)|11\rangle, & \epsilon(1-\epsilon)^2, \\
(\alpha|000\rangle+\beta|111\rangle)|10\rangle, & \epsilon(1-\epsilon)^2, \\
(\alpha|000\rangle+\beta|111\rangle)|11\rangle, & \epsilon(1-\epsilon)^2, \\
(\alpha|111\rangle+\beta|000\rangle)|01\rangle, & \epsilon^2(1-\epsilon), \\
(\alpha|111\rangle+\beta|000\rangle)|10\rangle, & \epsilon^2(1-\epsilon), \\
(\alpha|111\rangle+\beta|000\rangle)|11\rangle, & \epsilon^2(1-\epsilon), \\
(\alpha|111\rangle+\beta|000\rangle)|00\rangle, & \epsilon^3.
\end{array}$$

In order to recover the original $|\psi\rangle$, Bob applies the inverse operation of the encoding (9.3). As discussed in the introduction of this section, the encoding improves the transmission of quantum information if $\epsilon < 1/2$.

Remark 9.8. When Bob measures the syndrome, he does not learn anything about the quantum state (this is, the values of α and β). Thus, coherence is not destroyed. This has been possible because Alice has sent a many-qubit entangled state and Bob measured only collective properties of this state.

Remark 9.9. If we wish to use quantum-error correction repeatedly (e.g. for the stabilization of a state on a quantum computer; namely, a quantum memory, against decoherence from the environment) we must supply new ancillary qubits on every iteration. More precisely, erase them to the $|0\rangle$ state. According to Landauer's principle [**7, 8, 48**], erasure of information dissipates energy, so this process spends power.

9.3.2. Three-qubit phase-flip code. It is also possible to correct phase errors. This kind of error has no classical analogue. A phase-flip error is modeled through σ_z Pauli matrix, and affects the states of the computational basis as follows:

$$|0\rangle \to \sigma_z|0\rangle = |0\rangle, \quad |1\rangle \to \sigma_z|1\rangle = -|1\rangle.$$

Hence, a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is mapped into $\sigma_z|\psi\rangle = \alpha|0\rangle - \beta|1\rangle$.

Remark 9.10. The discussion in Section 4.1.1 may inspire us to observe that a phase-flip error in the computational basis $\{|0\rangle, |1\rangle\}$ becomes a bit-flip error in the basis $\{|+\rangle, |-\rangle\}$. Indeed, we have $\sigma_z|+\rangle = |-\rangle, \sigma_z|-\rangle = |+\rangle$.

By means of the Hadamard gate we may transform the vectors of the computational basis into new basis vectors and vice versa. Therefore, we use the encoding

$$|0\rangle \to |0_L\rangle = |+++\rangle, \quad |1\rangle \to |1_L\rangle = |---\rangle,$$

and use three bit-flip correcting code. We do the final decoding step by implementing the same array of gates as for the encoding, but in the reverse order.

9.3.3. Nine-qubit Shor code. The code we present in this section corrects the most general possible noise acting on a single qubit.

Notation 9.2. We will employ the following encoding:

$$|0\rangle \to |0_L\rangle \equiv \frac{1}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)$$

$$|1\rangle \to |1_L\rangle \equiv \frac{1}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)$$

So, a generic quantum state will be encoded in

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \to \alpha|0_L\rangle + \beta|1_L\rangle.$$

Fig. 9.3 shows the implementation of such encoding.

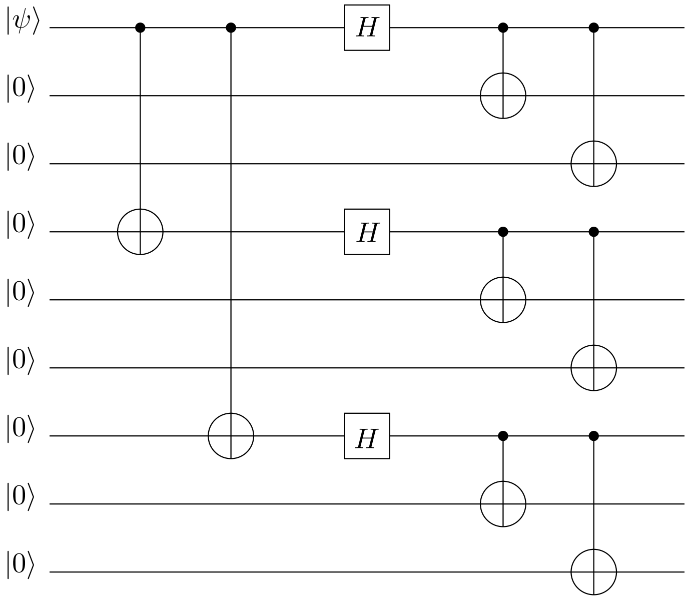

FIG. 9.3. 9-qubit Shor code encoding circuit. $|\psi\rangle$ is encoded into 9 qubits.

We shall now see how this code can correct both bit and phase-flip errors. Fig. 9.4 shows the quantum circuit for error recovery, which we shall now analyze: The method described in Section 9.3.1 can be applied to each three-qubit block since a single bit-flip error can be detected and corrected this way.

Let us now assume that it is a phase-flip error which has occurred in the first three-qubit block. Neglecting the normalization factor, the state of the first block is modified as

$$|000\rangle \pm |111\rangle \to |000\rangle \mp |111\rangle.$$

As observed in Remark 9.8, in order to detect this phase-flip error without disturbing the encoded quantum state $|\psi\rangle$, we must perform collective measurements.

We will measure the syndromes

$$y_0 = \sigma_x^{(1)}\sigma_x^{(2)}\sigma_x^{(3)}\sigma_x^{(4)}\sigma_x^{(5)}\sigma_x^{(6)}$$
$$y_1 = \sigma_x^{(1)}\sigma_x^{(2)}\sigma_x^{(3)}\sigma_x^{(7)}\sigma_x^{(8)}\sigma_x^{(9)}$$

This is, we will perform a parity check on the first and second blocks and on the first and third blocks in order to localize in which block the phase-flip has occurred.

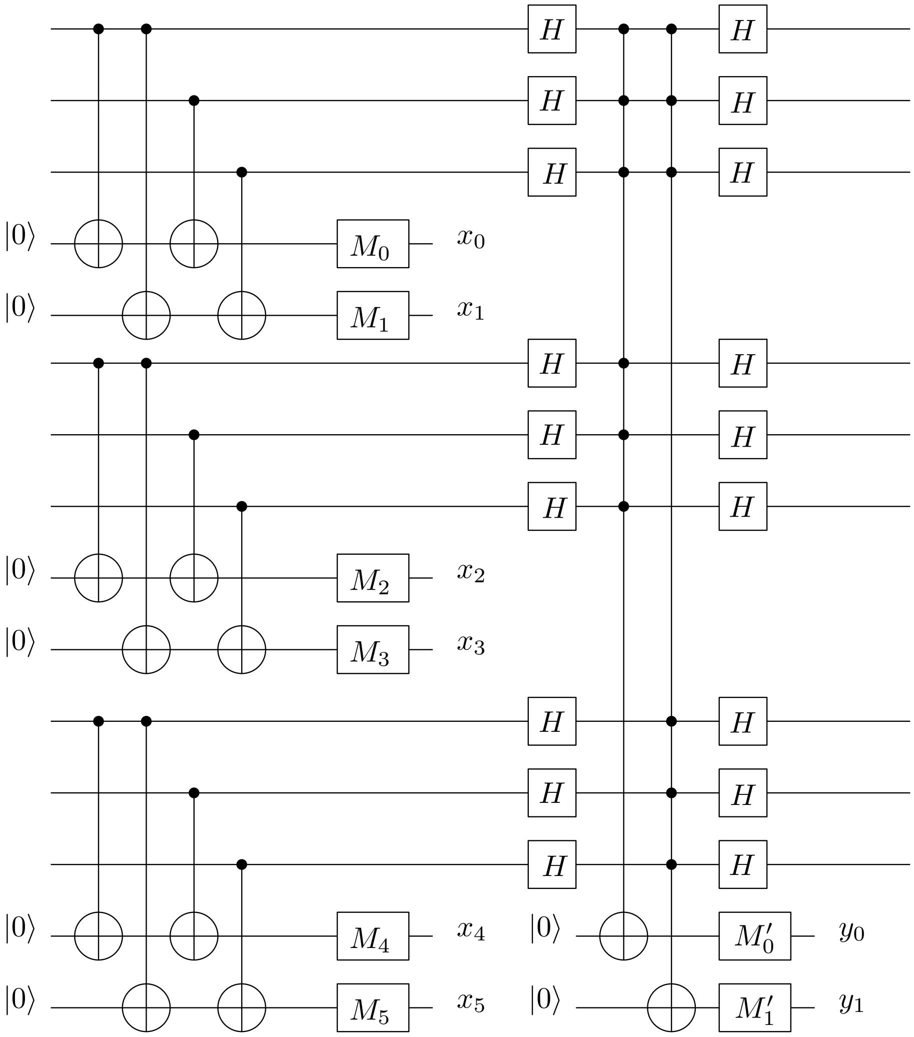

FIG. 9.4. 9-qubit Shor code syndrome-extraction circuit. Note that the three sub-blocks on the left are the same as in Fig. 9.2.

Therefore, $(y_0, y_1) = (-1, -1)$ means that the phase flip affects the first block of qubits. Similarly, $(1, -1), (-1, 1)$ and $(1, 1)$ correspond to third block, second block and no phase error, respectively.

To correct a phase error on block $i+1$ we apply the operator $\sigma_z^{(3i+1)}\sigma_z^{(3i+2)}\sigma_z^{(3i+3)}$, since it maps $|000\rangle \pm |111\rangle \to |000\rangle \mp |111\rangle$.

Remark 9.11. The nine-qubit Shor code does not only correct single-qubit bit and phase-flip errors. It also protects against arbitrary errors affecting a single qubit.

It is a fundamental feature of quantum error correction that a *continuum* of errors may be corrected by only correcting a discrete subset of them.

To understand this point, let us consider a single qubit interacting with the environment, which we can assume to be in a pure state $|0\rangle_E$ without loss of generality.

The most general evolution can be written as

$$U|0\rangle|0\rangle_E = |0\rangle|e_0\rangle_E + |1\rangle|e_1\rangle_E, \quad U|0\rangle|1\rangle_E = |0\rangle|e_2\rangle_E + |1\rangle|e_3\rangle_E.$$

Note that the states $|e_i\rangle_E$ need not be normalized or mutually orthogonal.

A simple calculation shows that, for a generic initial state of the system, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, one has

$$U|\psi\rangle|0\rangle_E = (\mathbb{I}_2|\psi\rangle)|e_I\rangle_E + (\sigma_z|\psi\rangle)|e_z\rangle_E + (\sigma_x|\psi\rangle)|e_x\rangle_E + (\sigma_x\sigma_z|\psi\rangle)|e_{xz}\rangle_E, \tag{9.4}$$

where we denoted

$$|e_I\rangle_E = \frac{1}{2}(|e_0\rangle_E + |e_3\rangle_E), \quad |e_z\rangle_E = \frac{1}{2}(|e_0\rangle_E - |e_3\rangle_E)$$
$$|e_x\rangle_E = \frac{1}{2}(|e_1\rangle_E + |e_2\rangle_E), \quad |e_{xz}\rangle_E = \frac{1}{2}(|e_1\rangle_E - |e_2\rangle_E).$$

We observe from that that the action of U can be expanded over the *discrete* set of operators (since $\sigma_y = i\sigma_x\sigma_z$):

$$\{\mathbb{I}_2, \sigma_x, \sigma_y, \sigma_z\},$$

which form a basis of the Hilbert space of 2 × 2 matrices.

When we measure the error syndrome, we are projecting the superposition $U|\psi\rangle|0\rangle_E$ onto one of the four states in which we expand it in (9.4).

In the case of Shor code, looking at the evolution of codewords $|0_L\rangle$ and $|1_L\rangle$, one obtains [**8**] that the final state of the environment is the same is initially either in the state $|0_L\rangle$ or $|1_L\rangle$. The deep reason that justifies this result is that $|0_L\rangle$ and $|1_L\rangle$ are entangled. Just by observing a single qubit (for any of them, the state would be maximally mixed $\rho = \frac{1}{2}\mathbb{I}_2$) it is impossible to tell them apart. Therefore, given an arbitrary $|\psi\rangle$, the environment cannot learn anything about α or β through the interaction with a single qubit, i.e., inducing single qubit errors. Since quantum information is not destroyed by this interaction, error recovery is possible.

9.4. CSS Codes

The codes we have seen so far exploit very little of classical codes. Here we present a quantum error correction code that corrects bit and phase flip errors independently, by using a quantum version of two linear codes. We will say that a code C is $[n,k]$ if it encodes a k-bit string into an n-bit string ($n > k$).

Definition 9.1. Let C_1 and C_2 be classical linear $[n,k_1],[n,k_2]$ codes, respectively, such that $C_2 \subset C_1$. For each codeword $\boldsymbol{x} \in C_1$, we define the quantum state

$$|x + C_2\rangle = \frac{1}{\sqrt{|C_2|}} \sum_{y \in C_2} |x + y\rangle,$$

where $|C_2|$ denotes the cardinality of C_2. The vector space spanned by $\{|x + C_2\rangle\}_{x\in C_1}$ defines a $[n, k_1 - k_2]$ quantum code, a Calderbank-Shor-Steane code, $CSS(C_1, C_2)$ for short.

From the definition we see that two different codewords $\boldsymbol{x}, \boldsymbol{x'}$ may lead to identical vectors $|x + C_2\rangle = |x' + C_2\rangle$. This will be the case if, and only if, $\boldsymbol{x} - \boldsymbol{x'} \in C_2$. Equivalently, if $\boldsymbol{x}$ and $\boldsymbol{x'}$ belong to the same coset of C_1/C_2. Otherwise, the states $|x + C_2\rangle$ and $|x' + C_2\rangle$ are orthogonal.

The dimension of the space $CSS(C_1, C_2)$ is $|C_1| / |C_2| = 2^{k_1-k_2}$ since it is the number of cosets of C_1/C_2; thus, $m = k_1 - k_2$ qubits can be encoded.

Error correction with CSS works as follows: Suppose that C_1 and $C_2^\perp$ can correct up to t errors[1]. Let us denote H_1 the parity check matrix for C_1 and H_2 that for $C_2^\perp$.

We define $\sigma_\alpha^s = \sigma_\alpha^{s_1} \otimes \sigma_\alpha^{s_2} \otimes \cdots \otimes \sigma_\alpha^{s_n}$, where $\alpha \in \{x, y, z\}$, $\sigma_\alpha^0 = \mathbb{I}_2$ and $s = (s_1, s_2, \ldots, s_n)$ is an n-bit vector.
By measuring σ_z^s for each row vector s of H_1 it can be shown [**14**] that one computes the syndrome for bit-flip errors. Similarly, the syndrome for phase-flip errors can be computed by measuring σ_x^r for each row vector r of H_2.
This way, arbitrary errors on any t qubits can be corrected. An important property of CSS codes is that error correction for phase errors and for bit flips is decoupled from each other [**14**].

Example 9.3. To illustrate the process described, let us consider the $[n = 7, k_1 = 4]$ Hamming code C_1 and the $[n = 7, k_2 = 3]$ code $C_2 = C_1^\perp$.

A generator matrix for C_1 is

$$G_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

and a parity check matrix is given by

$$H_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

It is trivial to check that $H_1 G_1 = 0$, the zero matrix, since the arithmetic we use is the one of $\mathbb{Z}_2$.

The code $C_2 = C_1^\perp$ has generator matrix $G_2 = H_1^T$ and parity check matrix $H_2 = G_1^T$. The code C_1 contains 16 codewords and the code C_2 contains 8. We list the latter as they will be needed later:

$$C_2 = \{0000000, 1010101, 0110011, 1100110, 0001111, 1011010, 0111100, 1101001\}.$$

In this case, we get a CSS code encoding $k_2 - k_1 = 1$ logical qubit into $n = 7$ physical qubits.

From Definition 9.1 we construct $|0_L\rangle$ and $|1_L\rangle$ as follows, taking $\boldsymbol{y_0} \in C_1$ and $\boldsymbol{y_{15}} \in C_1$ two representatives from different cosets of C_1/C_2; namely, $\boldsymbol{y_0} = 0000000, \boldsymbol{y_{15}} = 1111111$:

[1] If C is a linear $[n, k]$ code with generator matrix G and parity check matrix H, we define the dual code $C^\perp$ of C, which is the set of all codewords that are orthogonal to each codeword in C. The dual code $C^\perp$ is an $[n - k, n]$ code which is generated by H^T and has a parity check matrix G^T.

$$\begin{aligned}|0_L\rangle &= |y_0 + C_2\rangle \\ &= \frac{1}{\sqrt{8}}\left(|0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle\right. \\ &\quad \left. + |0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle\right).\end{aligned}$$

Similarly,

$$\begin{aligned}|1_L\rangle &= |y_{15} + C_2\rangle \\ &= \frac{1}{\sqrt{8}}\left(|1111111\rangle + |0101010\rangle + |1001100\rangle + |0011001\rangle\right. \\ &\quad \left. + |1110000\rangle + |0100101\rangle + |1000011\rangle + |0010110\rangle\right).\end{aligned}$$

We show now that, if the classical codes C_1 and $C_2^{\perp}$ correct up to t errors, the quantum CSS code can correct up to t qubits. As we have already seen in section 9.3.3, it suffices to correct bit-flip and phase-flip errors in order to correct arbitrary errors.

We can rewrite the state $|\tilde{v}\rangle = |v + C_2\rangle$ affected by amplitude and phase errors:

$$|\tilde{v}\rangle_{ap} = \frac{1}{\sqrt{2^{k_2}}} \sum_{w \in C_2} (-1)^{(v+w)\cdot e_p} |v + w + e_a\rangle,$$

where we denoted by $\cdot$ the bitwise scalar product. The n-bit vector $\boldsymbol{e}_a$ describes amplitude errors and the n-bit vector $\boldsymbol{e}_p$ describes phase errors. The errors are located in the positions where a 1 appears. Thus, we require that the weight (the number of ones) of these vectors is not greater than t.

In order to detect $\boldsymbol{e}_a$ we introduce a number of ancillary qubits, sufficient to store the error syndrome $|H_1 e_a\rangle$. This is, we map the state $|\tilde{v}\rangle_{ap}|0\rangle$ into

$$\frac{1}{\sqrt{2^{k_2}}} \sum_{w \in C_2} (-1)^{(v+w)\cdot e_p} |v + w + e_a\rangle |H_1 e_a\rangle.$$

The measure of the ancillary qubits will tell us where the bit-flip has occurred. Now, by applying a NOT gate to each of these qubits, we correct this kind of error; thus, obtaining the state

$$|\tilde{v}\rangle_p = \frac{1}{\sqrt{2^{k_2}}} \sum_{w \in C_2} (-1)^{(v+w)\cdot e_p} |v + w\rangle.$$

We have neglected the ancillary qubits, as they are needed no more. To correct the remaining phase errors, we will proceed, as usual, to turn them into amplitude errors and then reverse the transformation: Indeed, applying the Hadamard gate to each qubit, we obtain the state

$$\frac{1}{\sqrt{2^{k_2}}} \frac{1}{\sqrt{2^n}} \sum_{w \in C_2} \sum_{z=0}^{2^n - 1} (-1)^{(v+w)\cdot(e_p + z)} |z\rangle.$$

Changing the indexes $\boldsymbol{z'} = \boldsymbol{z} + \boldsymbol{e}_p$ ($\boldsymbol{z} = \boldsymbol{z'} + \boldsymbol{e}_p$, as the sum is the operation of $\mathbb{Z}_2$) we rewrite the state as

$$\frac{1}{\sqrt{2^{k_2}}} \frac{1}{\sqrt{2^n}} \sum_{w \in C_2} \sum_{z'=0}^{2^n - 1} (-1)^{(v+w)\cdot z'} |z + e_p\rangle.$$

Since, for a linear code $C[n,k]$, $\sum_{w\in C}(-1)^{w\cdot z} = 2^k$ if $z \in C^{\perp}$ and 0 otherwise [8], our state is

$$\frac{1}{\sqrt{2^{n-k_2}}} \sum_{z'\in C_2^{\perp}} (-1)^{v\cdot z'} |z' + e_p\rangle.$$

We have converted the phase error to an amplitude error. As we did previously, introducing ancillary qubits in the state $|0\rangle$ and applying the parity-check matrix $H_2^{\perp}$ for $C_2^{\perp}$ (which, in the case of the example, would be $H_1 = H_2^{\perp}$ since $C_2 = C_1^{\perp}$), we map $|z' + e_p\rangle|0\rangle$ into $|z' + e_p\rangle|H_2^{\perp} e_p\rangle$ and correct the error according to the measurement of the ancillary qubits. We finally obtain the state

$$\frac{1}{\sqrt{2^{n-k_2}}} \sum_{z'\in C_2^{\perp}} (-1)^{v\cdot z'} |z'\rangle$$

and apply the Hadamard gate to each qubit obtaining the original uncorrupted state $|\tilde{v}\rangle = |v + C_2\rangle$.

9.5. Fault-tolerant quantum computation

The discussion so far has assumed that encoding and decoding of quantum information can be achieved perfectly, as well as error recovery operations.

However, these are complex quantum operations which are subject to errors. Quantum logic gates which are present in quantum information processing may propagate errors in a quantum computer.

Quantum error correction has a double edge: The more qubits the system has, the more sensitivity it possesses to decoherence effects; however, to preserve the information on it, we do add extra qubits and perform quantum error correction.

The question which arises is: if we add a redundancy mechanism to protect a quantum computation, and a second redundancy mechanism to protect the first redundancy mechanism, and so on... does this process converge in a fault tolerant quantum computation?

In this section, we shall show that -under certain assumptions- an arbitrarily long quantum computation can be performed, given the noise in individual gates is below a critical threshold.

Example 9.4. Consider the CNOT gate. If a bit-flip error affects the control qubit ($|0\rangle \leftrightarrow |1\rangle$), then the error will also spread to the target qubit. For instance, the computation $CNOT(|0\rangle|0\rangle) = |0\rangle|0\rangle$ would turn into $CNOT(|1\rangle|0\rangle) = |1\rangle|1\rangle$.

Another example of error qubit propagation is known as backward sign propagation [7]: A phase error affecting the target qubit is also transferred, after the application of a CNOT gate, to the control qubit.

In order to implement a reliable quantum computation, fault-tolerant quantum gates must be applied. It is possible if we perform quantum logic operations directly on encoded states. For example, a fault tolerant CNOT gate could be implemented as shown in Fig. 9.5. It is easy to show [8] that if the CNOT gates are applied transversally (that is, bitwise), the truth table of the CNOT is verified for the logical qubits.

Remark 9.12. The CNOT gate is implemented fault-tolerantly, since each qubit in each code block is involved in a single gate. Therefore, errors can only propagate to at most one qubit in the other block, not inside the same block; thus making the CNOT gate fault-tolerant.

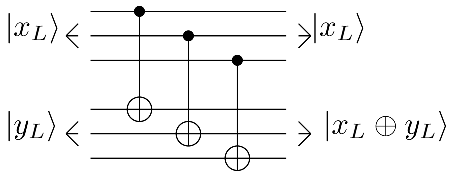

FIG. 9.5. A quantum circuit implementing a transversal CNOT gate between two logical qubits, encoded in three-qubit blocks.

9.5.1. The noise threshold for quantum computation. Given certain assumptions about the noise model (in the simplest case, random and uncorrelated errors) and provided the noise affecting individual quantum gates is below a certain threshold, the threshold theorem for quantum computation [**8**] tells us that it is in principle possible to efficiently implement arbitrarily long quantum computations.

The main idea behind this result is the use of *concatenated codes.*

Example 9.5. To grasp this concept, let us consider the following example, using the CSS code described in Example 9.3. It encodes a single logical qubit into a block of $n = 7$ qubits. In a concatenated code, each qubit of the block is itself a n-qubit block. This process is repeated until we reach L levels of concatenation, as shown in Fig. 9.6. Then, a single logical qubit is encoded into $n^L = 7^L$ physical qubits.

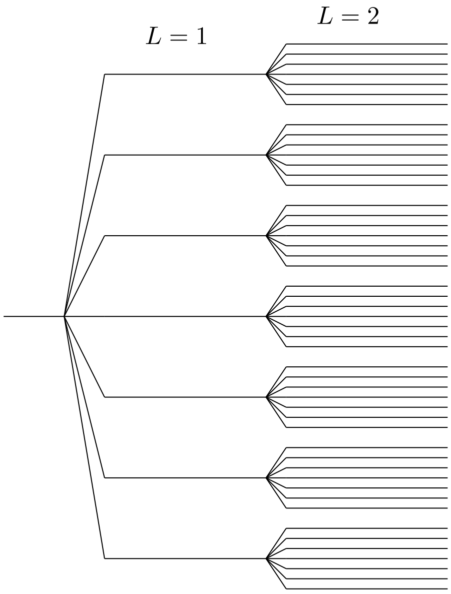

FIG. 9.6. Concatenation of a 7-qubit code up to $L =$ 2nd level.

Let us call ε the error probability per qubit and per unit of time (e.g., the time required to implement a single elementary quantum gate).

We shall call α the number of locations in the quantum circuit where an error may affect a single qubit (before that error correction is applied)[2].

At the first level of encoding ($L = 1$), error correction will fail if at least two qubits have been corrupted. Therefore, the probability of failure is

$$p_1 \approx c\varepsilon^2 \approx \alpha^2 \varepsilon,$$

where we have denoted $c \approx \alpha^2$ the number of ways in which a fault-tolerant circuit can introduce, at least, two errors.

At the second level of encoding ($L = 2$), we employ n^2 qubits and error correction will fail if -at least- two of e subblocks of size n fail. Hence, the failure probability is

$$p_2 \approx c p_1^2 \approx \alpha^2(\alpha^2 \varepsilon^2)^2.$$

Iterating this procedure, one finds that the failure probability at L-level concatenation is

$$p_L \approx c p_{L-1}^2 \approx \alpha^{2^{L+1}-2} \varepsilon^{2^L} = \frac{(\alpha^2\varepsilon)^{2^L}}{\alpha^2}.$$

Let us suppose that we wish to implement a computation of length T, where T is the number of *logic* quantum gates, with accuracy ε_0.

This can be approximated, for low values of ε_0 to require an error probability per logic gate $\leq \varepsilon_0/T$. Which is the number of levels of concatenation L we should apply?

The inequality

$$p_L \approx \frac{(\alpha^2\varepsilon)^{2^L}}{\alpha^2} \leq \frac{\varepsilon_0}{T}$$

is what we must match.

If $\varepsilon\alpha^2 \leq 1$, it can be rewritten as

$$L \geq \overline{L} \approx \log_2 \left(\frac{\log(T/\alpha^2\varepsilon_0)}{\log(1/\alpha^2\varepsilon)} \right).$$

Since $x^{\log y} = y^{\log x}$ (can be seen taking logarithms on each side), the total number of physical qubits

$$\overline{n}_{tot} \approx \left(\frac{\log(T/\alpha^2\varepsilon_0)}{\log(1/\alpha^2\varepsilon)} \right)^{\log_2 n}$$

grows only polylogarithmically with T and ε.

The condition $\varepsilon\alpha^2 \leq 1$ motivates the definition of the threshold error probability $\varepsilon_{th} = 1/\alpha^2$.

Remark 9.13. Note that the above result assume that the quantum computer hardware is capable of executing many quantum gates in parallel in a single time step. Otherwise, errors in concatenated codes would accumulate too quickly to allow for successful error correction.

[2] A typical value for α, in the case of codes correcting a single error (like Hamming's [7,4]) is $\alpha \sim 10^2$ [**8**].

Remark 9.14. For $\alpha \sim 10^2$, our result demands a noise threshold $\varepsilon_{th} \sim 10^{-4}$. In the literature [**8**], one can find various sophisticated calculations in which $\varepsilon_{th} \in (10^{-6}, 10^{-4})$. The numerical value of such threshold depends on the assumed features of the quantum computer hardware.

9.6. Degenerate codes

Another remarkable feature of quantum error correcting codes is that, in occasions, they can be used to correct more errors than those identified by the error syndrome [**20, 33, 61, 63**]. Such *degenerate* codes have been known since the 90's [**20, 61**], but have remained poorly understood since recently [**63**].

Shannon already discovered (Theorem 2.4) that the capacity of a noisy channel $\mathcal{N}$ with input symbol X and output symbol $Y = \mathcal{N}(X)$, its image under the action of the channel is

$$C = \sup_{p_X} \iota(X : \mathcal{N}(X)).$$

Moreover, the proof of the theorem (see [**66**]) shows that C is achieved by a random coding argument.

Remark 9.15. The maximization is over a single input to the channel; i.e., it does not require to consider the correlation of many inputs over many channel uses.

When one moves to the quantum analogue, he would expect that the quantum capacity is given by some measure of quantum correlations, maximized over the possible inputs. In order to show why it is not the case, we need some previous definitions.

Definition 9.2. It is convenient to introduce the concept of coherent information of a (bipartite) state:

$$I^c(\rho_{AB}) = S(\rho_B) - S(\rho_{AB}),$$

where S is the von Neumann entropy from Definition 7.1.

Definition 9.3. It is also convenient to introduce the concept of purification: If we are given a quantum state ρ_A describing a system Σ_A, it is possible to introduce another system Σ_B and define a *pure* state $|\phi_{AB}\rangle$ from $\Sigma_A + \Sigma_B$ such that it reduces to ρ_A when looking at system Σ_A alone:

$$Tr_B(|\phi_{AB}\rangle\langle\phi_{AB}|) = \rho_A.$$

This is always possible, as can be seen, for example, at [**48**]. More precisely, if $\rho_A = \sum_i p_i |i_A\rangle\langle i_A|$, we introduce Σ_B, having the same associated Hilbert space as Σ_A and we take an orthonormal basis $|i_B\rangle$. Then $|\phi_{AB}\rangle \equiv \sum\limits_i \sqrt{p_i}|i_A\rangle|i_B\rangle$.

We are in conditions to introduce the quantity Q_1, as the quantum analogue from Theorem 2.4, defined in [**63**]:

$$Q_1 \equiv \sup_\rho I^c(\mathcal{N}, \rho),$$

where

$$I^c(\mathcal{N}, \rho) = I^c(\mathbb{I} \otimes \mathcal{N}(|\phi_{AB}\rangle\langle\phi_{AB}|))$$

and $|\phi_{AB}\rangle$ is a purification of ρ.

The quantity Q_1 can be achieved using a random code on the typical subspace of the maximizing ρ, as one would expect. However, this rate has not always been optimal [**20, 61**]: If we consider the depolarizing channel introduced in Example 7.2, they show that codes with rates larger than Q_1 can be achieved, especially, when the depolarizing channel is very noisy (then Q_1 can be small or even zero [**63**]) this effect is more striking.

The correct formula is achieved when considering the *multi-symbol* case, rather the single-symbol we have already discussed.

Definition 9.4. The quantum capacity of a noisy channel $\mathcal{N}$ is given [**63**] by

$$Q \equiv \lim_{n\to\infty} \sup_{\rho_n} I^c(\mathcal{N}^{\otimes n}, \rho_n),$$

where the notation is self-explanatory.

Remark 9.16. When we take the limit $\lim_{n\to\infty}$, it means that we must consider the behavior of the channel on inputs which are entangled across many uses.

The difference between Q and Q_1 is closely related to the existence of *degenerate* quantum codes. Degeneracy is not a property of a quantum code alone, but a property of a code together with a family of errors it is designed to correct.

More precisely,

Definition 9.5. We say that a code $\mathcal{C}$ degenerately corrects a set of errors $\mathcal{E}_c$ if

- $\mathcal{C}$ corrects $\mathcal{E}_c$, and
- there are multiple errors in $\mathcal{E}_c$ mapped to the same error syndrome.

Definition 9.6. A code $\mathcal{C}$ is grossly degenerate if it has the further property that the number of collisions of errors to the same error syndrome is exponential in the code's block length.

In the case of the depolarizing channel, Q_1 is exactly the maximum rate achievable with a non-degenerate code. However, $Q > Q_1$, and this can be shown with the construction of a grossly degenerate code [**20, 61**].

In order to give a better understanding of how degenerate codes work, let us study quantum error correction in a more general framework (as suggested by [**63**]): when errors affecting n qubits occur.

Errors affecting n qubits can be described over a set of 4^n operators $\{E_k\}_k$ by means of Pauli operators, where

$$E_k \equiv \bigotimes_{j=1}^{n} \sigma_{i_j}^{(j)} : i_j \in \{0, x, y, z\}.$$

When considering the action of an arbitrary operator U on the n-qubit system $|\psi\rangle$ (plus the environment), the general expression is

$$U|\psi\rangle|0\rangle_E = \sum_{k=0}^{4^n-1} E_k|\psi\rangle|e_k\rangle_E.$$

The final environment states $|e_k\rangle_E$ need not concern us.

We have a set of possible errors, described by $\mathcal{E} = \{E_0, \dots, E_{4^n-1}\}$ and a subset $\mathcal{E}_c \subseteq \mathcal{E}$ of errors that can be corrected by the code.

The question we make to ourselves is what condition must be satisfied in order to make error correction possible?

It should be obvious that −given two different codewords, $|i_L\rangle$ and $|j_L\rangle$−, correctable errors should map them into orthogonal states in order to distinguish them with certainty (otherwise perfect error correction would be impossible):

$$\langle i_L | E_k^\dagger E_l | j_L \rangle = 0, \quad i \neq j \quad E_k, E_l \in \mathcal{E}_c.$$

Another necessary condition is that $\forall E_k, E_l \in \mathcal{E}_c$, we must not gain any information on the encoded state $|i_L\rangle$ from the measurement of the error syndrome. Otherwise, the quantum state would be disturbed (as discussed in the end of Section 9.3.3). This is the condition

$$\langle i_L | E_k^\dagger E_l | i_L \rangle = C_{kl},$$

where $C_{kl} \in \mathbb{C}$ does not depend on the state $|i_L\rangle$. Note that $C_{kl} = C_{lk}^*$.

The two previous conditions can be summarized in

$$\langle i_L | E_k^\dagger E_l | j_L \rangle = C_{kl} \delta_{i,j}, \tag{9.5}$$

where $E_k, E_l \in \mathcal{E}_c$ and the matrix $C = (C_{kl})_{kl}$ is Hermitian.

Proposition 9.1. Error correction is possible if, and only if, (9.5) is satisfied.

The proof of this proposition can be found at [**53**].

With Eq. (9.5) we arrive at an equivalent definition, which is more operative, of degeneracy [**8**]:

Definition 9.7. A code $\mathcal{C}$ is non-degenerate if

$$C_{kl} = \delta_{k,l}, \quad \forall\ k, l.$$

Otherwise, if $C_{kl} \neq \delta_{k,l}$ for some C_{kl}, the code is degenerate.

We have already seen examples of degenerate and non-degenerate codes, which we shall explain below:

Example 9.6. The three-qubit bit-flip code from Section 9.3.1 is non-degenerate.
Indeed, let us check that (9.5) is fulfilled: We have 3 correctable errors, which are

$$\begin{aligned} E_1 &= \sigma_x^{(1)} \otimes \sigma_0^{(2)} \otimes \sigma_0^{(3)} = E_1^\dagger, \\ E_2 &= \sigma_0^{(1)} \otimes \sigma_x^{(2)} \otimes \sigma_0^{(3)} = E_2^\dagger, \\ E_3 &= \sigma_0^{(1)} \otimes \sigma_0^{(2)} \otimes \sigma_x^{(3)} = E_3^\dagger, \end{aligned}$$

where $\sigma_0^{(j)} = \mathbb{I}_2^{(j)}$.

Since Pauli matrix σ_x is $\sigma_x^2 = \mathbb{I}_2$, $E_1^\dagger E_1 = E_2^\dagger E_2 = E_3^\dagger E_3 = \mathbb{I}_2$ obtaining

$$\langle i_L | E_k^\dagger E_k | j_L \rangle = \delta_{i,j}.$$

Remaining cases are easy to check. For example

$$\langle 0_L | E_2^\dagger E_3 | 0_L \rangle = \langle 000 | \sigma_0^{(1)} \otimes \sigma_x^{(2)} \otimes \sigma_x^{(3)} | 000 \rangle = \langle 000 | 011 \rangle = 0.$$

Similarly, for $|1_L\rangle$. In summary,

$$\langle i_L|E_k^\dagger E_l|i_L\rangle = 0, \quad k \neq l.$$

This means that Eq. (9.5) holds with $C_{kl} = \delta_{k,l}$. Hence, the code is non-degenerate.

Example 9.7. The nine-qubit Shor code introduced in Section 9.3.3 is degenerate.
Indeed, let us see that C is nor the identity matrix anymore: Let us consider the errors E_1 and E_2, corresponding to phase-flip errors in qubits 1 and 2 respectively: $E_1 = \sigma_z^{(1)} \otimes \sigma_0^{(2)} \otimes \cdots \sigma_0^{(9)}$, and $E_2 = \sigma_0^{(1)} \otimes \sigma_z^{(2)} \otimes \cdots \sigma_0^{(9)}$.
The effect of E_1 and E_2 is the same onto the logical qubits $|0_L\rangle$ and $|1_L\rangle$:

$$|0_L^{(1)}\rangle = E_1|0_L\rangle = \frac{1}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle),$$
$$|1_L^{(1)}\rangle = E_1|1_L\rangle = \frac{1}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle);$$

whereas

$$|0_L^{(2)}\rangle = E_2|0_L\rangle = \frac{1}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle),$$
$$|1_L^{(2)}\rangle = E_2|1_L\rangle = \frac{1}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle).$$

Hence, we obtained $|0_L^{(1)}\rangle = |0_L^{(2)}\rangle$ and $|1_L^{(1)}\rangle = |1_L^{(2)}\rangle$. This means that

$$\langle i_L|E_1^\dagger E_2|i_L\rangle = 1, \quad i = 0, 1 \Rightarrow C_{12} = 1 \neq \delta_{1,2} = 0.$$

Thus, the nine-qubit Shor code is a case of a degenerate code.

Remark 9.17. Observe that the states $|0_L^{(j)}\rangle$ from the previous example, for $j \in \{1, 2, 3\}$ would be indistinguishable, yet the code is capable of correcting the error. The error syndrome is the same for all cases. This is the key feature of degeneracy which allows to improve the quantity $Q > Q_1$.

Part 4

Physical Implementations

Chapter 10
Physical Realizations

In this last chapter, we describe which physical implementations are being developed nowadays, and how they are accomplished. We consider various types of technologies (Ion traps, cavity quantum electrodynamics, photonic realizations...) and analyze their benefits and drawbacks. We focus on the photonic realization of entanglement via sporadic parametric down conversion and the optical realization of protocols and computation. In particular, we treat the Cirac-Zoller CNOT gate and the Jaynes-Cummings model [**26, 43, 48, 8**].

10.1. Photonic Realizations

At present, the only appropriate system for long-distance communication of quantum states is the photon, since photons can travel long distances with low loss (in optical fibers or in free space). The state of a single photon can be manipulated by means of basic linear optical components, such as phase shifters and beam splitters, as we shall discuss in this section.

Definition 10.1. An optical component is said to be linear if its output modes (with creation and annihilation operators $\hat{b}_j^\dagger$ and $\hat{b}_j$) are a linear combination of its input modes (with creation and annihilation operators $\hat{a}_j^\dagger$ and $\hat{a}_j$).

$$\hat{b}_j^\dagger = \sum_k M_{jk} \hat{a}_k^\dagger.$$

10.1.1. The phase shifter.

Definition 10.2. A phase shifter of phase ϕ on the Fock state $|m\rangle$ is defined by the transformation U_P:

$$U_P(\phi) = e^{i\phi\hat{m}} = e^{i\phi\hat{a}^\dagger\hat{a}}.$$

Therefore, the Fock state $|m\rangle$ is mapped into $e^{i\phi m}|m\rangle$.

The phase shifter is implemented in practice with a slab of transparent medium with refractive index n; with $n \neq n_0$, where n_0 is the refractive index of free space. Hence, the wave vector in the medium is

$$k = n\omega/c$$

and in free space

$$k_0 = n_0\omega/c,$$

where $\omega/2\pi$ is the photon frequency and c is the speed of light in vacuum.

If the photon travels a distance L through the medium, its phase will change by e^{ikL}, which is different from the phase change e^{ik_0L} it would have undergone when traveling the same distance in free space.

Thus, the phase shift from the definition would be $\phi = kL$ or $\phi = k_0L$, depending on the medium.

In Fig. 10.1 (left) we have represented two phase shifters and a beam splitter.

10.1.2. The beam splitter.

Definition 10.3. The beam splitter acts on two modes through the unitary transformation U_B:

$$U_B(\theta, \phi) = \begin{pmatrix} \cos\theta & -e^{i\phi}\sin\theta \\ e^{-i\phi}\sin\theta & \cos\theta \end{pmatrix},$$

where the relation between the input and output modes is the linear mapping [**8**]

$$\hat{a}_l^\dagger|0\rangle \mapsto \sum_{m=1,2}(U_B)_{ml}\hat{b}_m^\dagger|0\rangle.$$

In particular, for the input state

$$|mn\rangle = \frac{(\hat{a}_1^\dagger)^m}{\sqrt{m!}}\frac{(\hat{a}_2^\dagger)^n}{\sqrt{n!}}|00\rangle,$$

and according to (5.10), we obtain the output state

$$\begin{aligned} U_B|mn\rangle &= \frac{1}{\sqrt{m!n!}}\left(\sum_{i=1}^{2}(U_B)_{i1}\hat{b}_i^\dagger\right)^m \left(\sum_{j=1}^{2}(U_B)_{j2}\hat{b}_j^\dagger\right)^n |00\rangle \\ &= \frac{1}{\sqrt{m!n!}}(\cos\theta\ \hat{b}_1^\dagger + e^{-i\phi}\sin\theta\ \hat{b}_2^\dagger)^m(-e^{i\phi}\sin\theta\ \hat{b}_1^\dagger + \cos\theta\ \hat{b}_2^\dagger)^n|00\rangle. \end{aligned}$$

Example 10.1. In the dual-rail representation, a single photon can follow two different paths, and the states of the qubit $|0\rangle, |1\rangle$ correspond to the photon following one path or the other.

The two logical states can be written as $|0\rangle = \hat{a}_0^\dagger|0\rangle_0|0\rangle_1 = |1\rangle_0|0\rangle_1$ and $|1\rangle = \hat{a}_1^\dagger|0\rangle_0|0\rangle_1 = |0\rangle_0|1\rangle_1$, where the operators $\hat{a}_0^\dagger$ and $\hat{a}_1^\dagger$ create a photon in the input modes 0 and 1; $|0\rangle_0$ and $|0\rangle_1$ are the vacuum states corresponding to these modes.

As seen in Fig. 10.1, a beam splitter

$$U_B(\theta = \pi/4, \phi = -\pi/2) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

implements the transformation

$$|0\rangle \mapsto \frac{1}{\sqrt{2}}(|0'\rangle + i|1'\rangle), \quad |1\rangle \mapsto \frac{1}{\sqrt{2}}(i|0'\rangle + |1'\rangle),$$

where we have represented the states $|0'\rangle = \hat{b}_{0'}^\dagger|0\rangle_{0'}|0\rangle_{1'} = |1\rangle_{0'}|0\rangle_{1'}$ and $|1'\rangle = \hat{b}_{1'}^\dagger|0\rangle_{0'}|0\rangle_{1'} = |0\rangle_{0'}|1\rangle_{1'}$. Here, the operators $\hat{b}_{0'}^\dagger$ and $\hat{b}_{1'}^\dagger$ create a photon in the output modes $0'$ and $1'$. In Fig. 10.1 we see that this beam splitter, together with two $-\pi/2$ phase shifters, implements a Hadamard gate:

Indeed, the sequence of gates

$$U_P(\phi=-\pi/2)U_B(\theta=\pi/4,\phi=-\pi/2)U_P(\phi=-\pi/2)$$

transforms the input states $|0\rangle$ and $|1\rangle$ as follows (with an obvious abuse of notation):

$$\begin{aligned}
|0\rangle=|1\rangle_0|0\rangle_1 &\rightarrow |1\rangle_0|0\rangle_1\rightarrow\frac{1}{\sqrt{2}}(|1\rangle_0|0\rangle_1+i|0\rangle_0|1\rangle_1)\\
&\rightarrow \frac{1}{\sqrt{2}}(|1\rangle_0|0\rangle_1+|0\rangle_0|1\rangle_1)=\frac{1}{\sqrt{2}}(|0'\rangle+|1'\rangle).\\
|1\rangle=|0\rangle_1|1\rangle_0 &\rightarrow -i|0\rangle_0|1\rangle_1\rightarrow\frac{-i}{\sqrt{2}}(i|1\rangle_0|0\rangle_1+|0\rangle_0|1\rangle_1)\\
&\rightarrow \frac{1}{\sqrt{2}}(|1\rangle_0|0\rangle_1-|0\rangle_0|1\rangle_1)=\frac{1}{\sqrt{2}}(|0'\rangle-|1'\rangle), \qquad (10.1)
\end{aligned}$$

which is the precise behavior of a Hadamard gate.

Example 10.2. We can also represent a qubit by its polarization: the two polarization states $|h\rangle, |v\rangle$ (horizontal, vertical) stand for the states $|0\rangle, |1\rangle$.

In Fig. 10.1 we show how the CNOT gate can be implemented −up to a sign factor− provided the dual-rail qubit is the control and the polarization qubit the target. We introduce a polarization rotator ($R : |h\rangle \rightarrow |v\rangle, |v\rangle \rightarrow -|h\rangle$) in the upper $1'$ path.

Indeed, we have

$$\begin{aligned}
|0\rangle|h\rangle &= |1\rangle_0|0\rangle_1|h\rangle\rightarrow|1\rangle_{0'}|0\rangle_{1'}|h\rangle=|0'\rangle|h\rangle,\\
|0\rangle|v\rangle &= |1\rangle_0|0\rangle_1|v\rangle\rightarrow|1\rangle_{0'}|0\rangle_{1'}|v\rangle=|0'\rangle|v\rangle,\\
|1\rangle|h\rangle &= |0\rangle_0|1\rangle_1|h\rangle\rightarrow|0\rangle_{0'}|1\rangle_{1'}|v\rangle=|1'\rangle|v\rangle,\\
|1\rangle|v\rangle &= |0\rangle_0|1\rangle_1|v\rangle\rightarrow-|0\rangle_{0'}|1\rangle_{1'}|h\rangle=-|1'\rangle|h\rangle.
\end{aligned}$$

This is exactly the behaviour of a CNOT gate.

10.2. Cavity quantum electrodynamics

Cavity quantum electrodynamics (CQED) stands for a set of techniques enabling the interaction of single atoms and single photons inside a resonating cavity.

In this section we will focus on experiments realized with atoms whose valence electrons are in states with a very large principal quantum number n; they are called *Rydberg atoms*. More precisely, we consider alkali atoms, which have a single valence electron, very far from the atomic nucleus; therefore, its electric dipole moment is very high and can be used to achieve the so-called *strong-coupling regime*. This means that the coherent evolution of a single atom coupled to a single photon −stored in a high quality cavity− overwhelms the incoherent dissipative processes[1]. Thus, atom-photon entanglement can be produced before decoherence dominates.

[1] Quality is a measure of the rate at which a vibrating system dissipates its energy. We define the quality factor Q as 2π times the ratio of the stored energy over the energy lost per cycle. Typical values are $Q \sim 3\cdot 10^8$ [**8**].

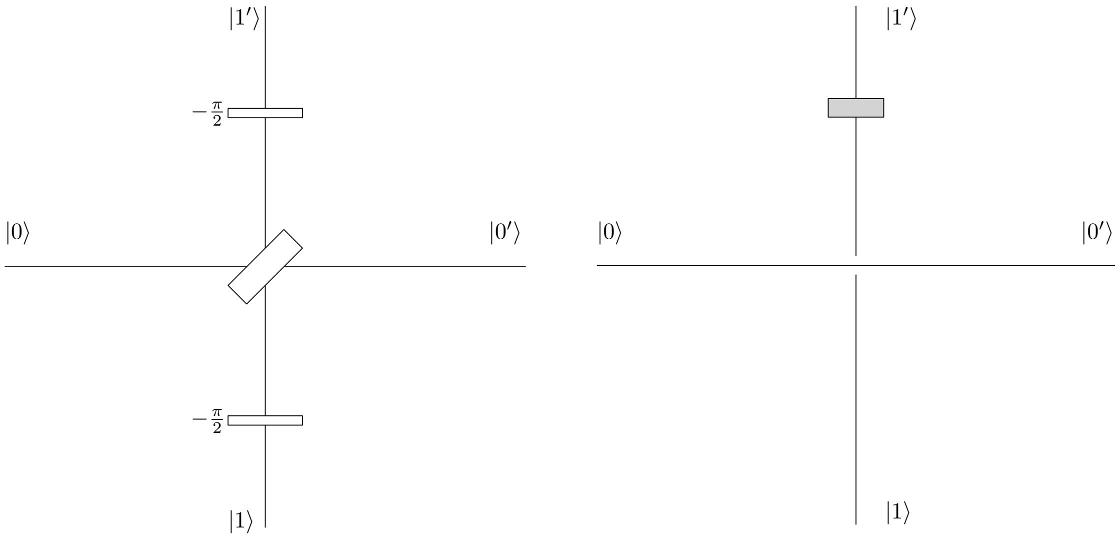

FIG. 10.1. Optical scheme of the Hadamard gate (left) and CNOT gate (right).

An important observation is that the energy separation $E_n - E_{n-1}$ between two consecutive atomic levels is very low: It corresponds to frequencies in the range $10 \sim 50$GHz, whereas when $n \sim 1$ they would correspond to optical frequencies in the order of 10^{15}Hz.

Note that these radio frequencies are available in laboratories, so that resonant cavities can be excited and then used to manipulate atoms. Also, the lifetime of Rydberg atoms is very long [**8**]. For example, for $n \sim 50$, and a high angular momentum $l \sim n$, the transition frequency between states with principal quantum numbers n and $n-1$ is in the microwave range and its lifetime is about 30ms. The microwave wavelength is in the order of centimeters, so it is very convenient for experimental manipulation.

In Fig. 10.2, we have shown the typical apparatus setup for CQED experiments.

- Alkali atoms leave the oven O and are excited to the appropriate Rydberg state by means of appropriately tuned laser pulses L. In order to select atoms with well defined velocity, Doppler effect is used. Although the source emits atoms randomly, pulsed lasers allow to prepare the Rydberg states in $O(\mu\text{s})$. This means that the position of each atom flying inside the apparatus is known with $O(\text{mm})$ precision, allowing to address and control individual atoms.
- The prepared Rydberg atom crosses one or more (usually microwave superconducting) cavities, R_1, C, R_2, resonant with the transition between two atomic levels $|g\rangle, |e\rangle$. The relaxation time of the fields applied in R_1 and R_2 is $O(\text{ns})$ and therefore they do not produce any entanglement between the atom and the microwave radiation field. The so-called Rabi oscillations produced in these cavities are of the order of 10μs, much longer than the relaxation time. These cavities are used for preparing the initial state in the desired superposition $\alpha|g\rangle + \beta|e\rangle$.
  The cavity C is prepared in the vacuum state $|0\rangle$ with no photons (the photon mean number can be reduced to 10^{-1}); it can evolve to the one-photon state $|1\rangle$ after the interaction with the atom. The photon storage time is in the order of $O(\text{ms})$, which is much larger than the atom-cavity interaction time, a few tens of μs.

- Finally, the up/down state of the atom is measured using two detectors D_g and D_e. This is accomplished by means of ionizing the atom with a static electric field $O(10^2\text{V/cm})$ and detecting the resulting electron. The detectors D_g and D_e are very selective and will only ionize the atom if it is in the state $|g\rangle$ or $|e\rangle$, respectively. For example, circular Rydberg states for rubidium atoms with $n = 49, 50, 51$ can be distinguished with this technique.

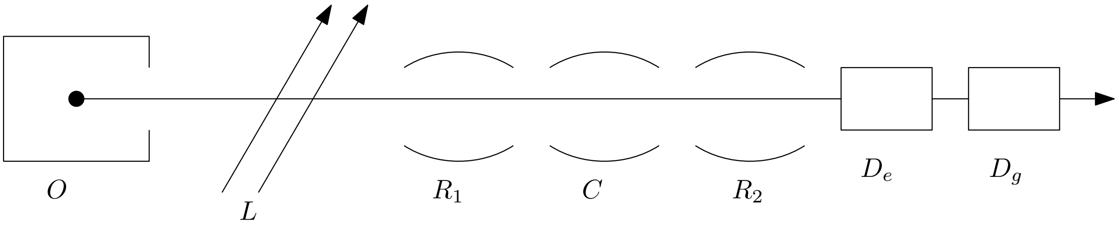

FIG. 10.2. A sketch of a cavity quantum electrodynamics apparatus: the atoms leave the oven O and are excited into the desired Rydberg state by pulsed lasers L. They enter the cavities R_1, C, R_2 and are finally detected using state selective field ionization in D_e and D_g.

Remark 10.1. The use of R_2 is to be able to fully determine the (12) parameters that define the quantum operation $\rho \overset{C}{\mapsto} \rho'$. Also, the fields in R_1 and R_2 can be considered classical. When an atom interacts with a classical field, its state remains pure [**8**].

10.3. The Jaynes-Cummings model

We now revisit the atom-field interaction, but this time the field is also quantized, as we described in Section 5.2.1. More precisely, we consider the interaction of a two-level atom with a single mode of the quantized electromagnetic field. The state vector of the atom, at time t has the form

$$|\psi_{atom}\rangle = c_e(t)|e\rangle + c_g(t)|g\rangle,$$

where we have followed the notation of Section 10.2. The state of the field, in the Fock basis (see Section 5.2.2), is

$$|\psi_{field}\rangle = \sum_n c_n(t)|n\rangle.$$

Thus, the collective atom-field state is the tensor product of both,

$$|\psi_{atom-field}\rangle = \sum_n c_{n,e}(t)|e\rangle \otimes |n\rangle + c_{n,g}(t)|g\rangle \otimes |n\rangle,$$

which is, in general, an entangled state.

Definition 10.4. The total atom-field Jaynes-Cummings Hamiltonian is written as

$$H = H_0^{atom} + H_0^{field} + V,$$

where

$$\begin{aligned} H_0^{atom} &= \hbar(E_e|e\rangle\langle e| + E_g|g\rangle\langle g|), \\ H_0^{field} &= \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right), \\ V &= \hbar\tilde{g}(\hat{a} + \hat{a}^\dagger)(\hat{\sigma}_- + \hat{\sigma}_+). \end{aligned}$$

Note that the Hamiltonian corresponding to the field is the same as the quantum harmonic oscillator (5.4). Here we use the lowering and raising operators $\hat{\sigma}_-$ and $\hat{\sigma}_+$, which are linear combinations of the Pauli matrices:

$$\sigma_+ = |e\rangle\langle g| = \left(\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right) = \frac{1}{2}(\sigma_x + i\sigma_y),$$

$$\sigma_- = |g\rangle\langle e| = \left(\begin{array}{cc} 0 & 0 \\ 1 & 0 \end{array}\right) = \frac{1}{2}(\sigma_x - i\sigma_y) = \sigma_+^\dagger.$$

Finally, $\tilde{g}$ is the so-called atom-field coupling constant [**43**].

Let us now analyze the terms of this Hamiltonian: The terms $(\hat{a} + \hat{a}^\dagger)$ and $(\hat{\sigma}_- + \hat{\sigma}_+)$ represent the time-dependent part of the driving field, and the induced transitions $|e\rangle \to |g\rangle$ and $|g\rangle \to |e\rangle$, respectively. We have the following operator products:

- $\hat{a}\hat{\sigma}_-$: This mediates the transition $|e\rangle \to |g\rangle$ and the absorption of a photon. It corresponds to a total energy loss of $\sim 2\hbar\omega$.
- $\hat{a}\hat{\sigma}_+$: This corresponds to the transition $|g\rangle \to |e\rangle$ and the absorption of a photon. This process is called stimulated absorption, and the total energy is conserved.
- $\hat{a}^\dagger\hat{\sigma}_-$: This governs the transition $|e\rangle \to |g\rangle$, together with the emission of a photon. This process is known as stimulated emission, and conserves the total energy.
- $\hat{a}^\dagger\hat{\sigma}_+$: This describes the emission of a photon and the transition $|g\rangle \to |e\rangle$. It has an energy gain of $\sim 2\hbar\omega$.

10.3.1. Atom-atom entanglement in CQED. This is one of the first demonstrations of creating controlled entanglement of two qubits [**27**] using the apparatus described in Fig. 10.2.

The idea is the following: The two qubits correspond to two-level atoms, which interact, one *after* the other, with a single mode quantized field, which is resonant with the atomic transition frequency; hence, described by the Jaynes-Cummings interaction.

We arrange things in order that the first atom will exit the cavity C in a balanced coherent superposition of the $|e\rangle$ and $|g\rangle$ states. So, the mode will be in a balanced superposition of $|n\rangle$ and $|n+1\rangle$ photons. Hence, the first atom and the field will be in a maximally entangled state.

The second atom will then recover this information from the field, left there by the first atom and, once done, the two atoms will be maximally entangled, whereas the field will remain totally disentangled.

This is a quite generic scenario, in which the atoms do not interact directly with each other; they interact through the field, as if it were some kind of data bus or an intermediate information storage.

More precisely,

- The atom$_1$-atom$_2$-field system is initially prepared in the state
$$|\psi\rangle = |e_1\rangle \otimes |g_2\rangle \otimes |0\rangle.$$
- The velocity of atom$_1$ is such that the interaction time t_1 with the field mode satisfies $\Omega t_1 = \pi/2$, where $\Omega = 2\tilde{g}\sqrt{n+1}|_{n=0}$ is the vacuum Rabi frequency (the frequency which defines and confines the field mode).

- The total state after the atom leaves the cavity is
$$|\psi'\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle \otimes |g_2\rangle \otimes |0\rangle - i|g_1\rangle \otimes |g_2\rangle \otimes |1\rangle).$$
- Choosing the velocity of the second atom such that $t_2 = 2t_1$ (therefore, $\Omega t_2 = \pi$), then it will leave the cavity in $|d_2\rangle$, provided the first atom left the field mode empty; or it will leave the cavity in $|e_2\rangle$ −with unit probability−, if the first atom deposited a photon in the field. Thus, the final state reads
$$|\psi''\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle \otimes |g_2\rangle \otimes |0\rangle - |g_1\rangle \otimes |e_2\rangle \otimes |0\rangle),$$
which can be rewritten as
$$|\psi''\rangle = \frac{1}{\sqrt{2}}(|e_1\rangle \otimes |g_2\rangle - |g_1\rangle \otimes |e_2\rangle) \otimes |0\rangle.$$

Remark 10.2. Note that $|\psi''\rangle$ is maximally entangled in the atomic states, whereas it is separable in the atoms and field subsystems.

10.4. Ion Traps

The main idea of this method is to have a string of ions trapped in well controlled positions and to individually address each ion with the use of laser pulses. Due to progress in laser technology, the degree of control over the states of trapped ions is continuously increasing and generation and coherent manipulation of entangled states has been achieved; e.g., up to 14 qubits [**45**] in a linear ion trap.

We shall describe a particular case of a trap, which is the Paul trap: the ions are confined by a spatially varying time-dependent radiofrequency field oscillating at frequency ω_{RF}. The ions in the trap repel each other through Coulomb forces; therefore, they undergo collective motion in their translational degree of freedom (along the trap axis, which is weakly confined by a shallow harmonic potential).

If ω_x, ω_y and ω_z are the frequencies along the three principal axes of the trap (let us suppose it is z-oriented), the field is such that the trapping frequency $\omega_t \equiv \omega_z \ll \omega_x, \omega_y$, so we can limit our considerations of motion along the z axis. In Fig. 10.3 we can see a photo [**11**] of a linear ion trap.

Typical values used in experiments [**8**] are $\omega_{RF} \sim 50$MHz, with applied voltages from 100V to 500V, trap sizes of $\sim$ 1mm. This leads to an harmonic motion of the ion in the z-direction of frequency $\omega_t/2\pi$ from $\sim$ 1MHz to $\sim$ 5MHz.

10.5. The Cirac-Zoller CNOT gate

10.5.1. Experimental results. To conclude this chapter, we present the implementation of the Cirac-Zoller CNOT quantum gate. Its experimental scheme, suggested in [**15**] and sketched in Fig. 10.4 and with its image in Fig. 10.5, with the trapped ions technique, which was realized in [**10, 11**], with its experimentally observed truth table shown in Fig. 10.6 and its numerical

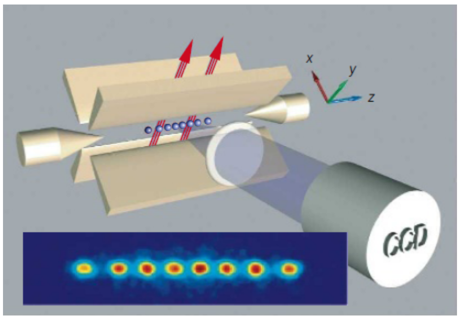

FIG. 10.3. A linear quadrupole Paul trap (beige) containing individually addressed $^{40}\mathrm{Ca}^{+}$ ions (blue). After cooling by laser beams (red) the trapped ions form a string and are imaged using a charge-coupled device (CCD). In the CCD image, the spacing between ions is $\sim 8\mu$m.

results (reproduced from [**10**]):

	$\|gg\rangle$	$\|ge\rangle$	$\|eg\rangle$	$\|ee\rangle$
$\|gg\rangle$	0.74(3)	0.13(3)	0.05(3)	0.08(3)
$\|ge\rangle$	0.15(3)	0.71(5)	0.06(1)	0.08(2)
$\|eg\rangle$	0.01(2)	0.08(3)	0.14(4)	0.77(3)
$\|ee\rangle$	0.03(3)	0.02(1)	0.72(6)	0.22(4)

In the implementation of [**10**], two $^{40}\mathrm{Ca}^{+}$ ions are held in a linear Paul trap. The pulse sequence for its operation requires $\sim$ 500μs (and the decoherence time scale is of the order of 1ms). In the experiment, the fidelity of the gate is 71%. The principal sources of errors are laser frequency noise and heating, due to stochastically fluctuating electric fields. Other sources of errors are spontaneous decay of the ion within the detection time and spurious fluorescence from the adjacent ion.

Remark 10.3. An obvious observation from the experiment we described is that the ions we want to interact with need to be adjacent. Is it still possible to implement a $\mathrm{CNOT}_{i,j}$ in a ion trap containing n ions, where we use the ith ion as control and the jth ion as target? In [**29**] this is achieved for $n = 4$, by means of transporting the qubits, and rearranging them in such a way that the desired interaction is achieved. The transport is controlled by time-varying potentials and other zones of the processor are dedicated to qubit storage (a quantum memory).

10.5.2. Theoretical approach.

Notation 10.1. Let us now focus on the theoretical model introduced in [**15**]. The symbols e and g will be used as labels for the ionic excited and ground states, respectively. There is the need to distinguish between the degenerate sub-level of the excited state, so we will use the suffix $q = 0, 1$. Finally, we will refer to ion $n \in \{1, \ldots, N\}$ with an additional suffix.

The dynamics of ion n is described by the Jaynes-Cummings Hamiltonian (assuming the Lamb-Dicke approximation) [**43**]:

$$H_{n,q} = \frac{\Omega_{\mathrm{eff}}}{2}\left(|e_q\rangle_n\langle g|\hat{a} + |g\rangle_n\langle e_q|\hat{a}^{\dagger}\right),$$

where Ω_{eff} is the effective Rabi frequency.

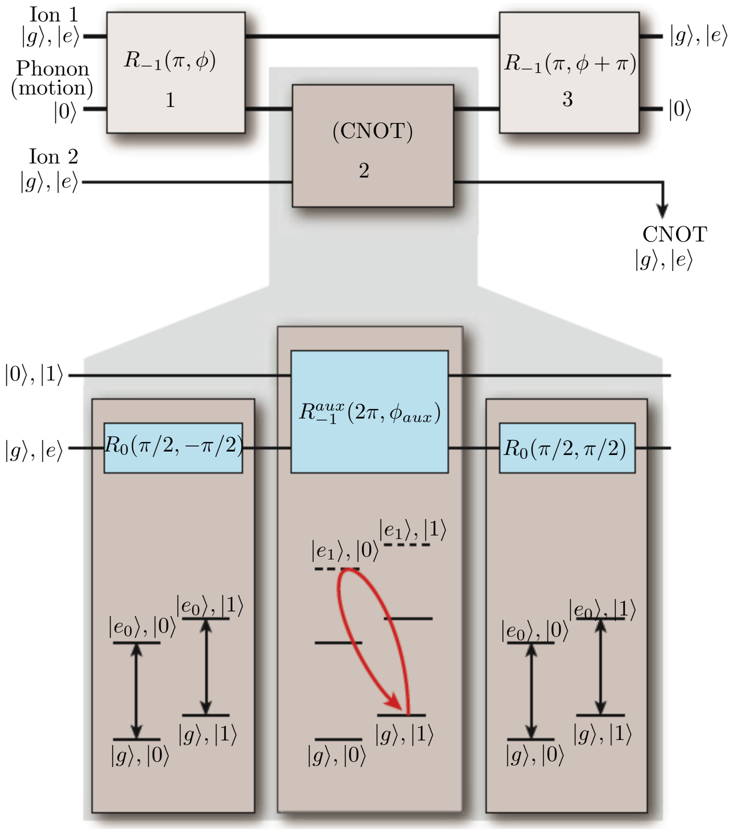

FIG. 10.4. Scheme for the realization of experimental CNOT gate described in [**10**]: Two ions in the same trap are initially prepared in their motional ground state. In Step 1, a low-sideband laser pulse $R_{-1}(\pi, \phi)$ is applied to the first ion (the control qubit). This maps the excited state amplitude to the first excited state of the motional mode (SWAP operation). In Step 2, a CNOT operation is performed between the motion qubit and the spin state of ion 2. Finally, in Step 3, Step 1 is reversed.

Similarly to Section 10.3.1, in which excitations of the radiation field (photons) were created and used to carry quantum information, here $\hat{a}$ and $\hat{a}^\dagger$ annihilate and create, respectively, excitations of the collective center-of-mass motion in the translational degree of freedom (phonons).

Thus, the energy is exchanged between the electronic degrees of freedom of the ions and the single quantized mode of their collective translational dynamics, with characteristic frequency Ω_{eff} .

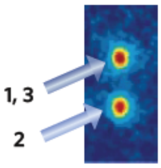

FIG. 10.5. CCD image of the two ions forming the CNOT gate in [**10**] experiment, published in [**11**]. Following the process of Fig. 10.4 a laser pulse is first applied to ion 1 and its state becomes entangled with the field mode of the motion phonon; then a sequence of pulses performs a CNOT operation to ion 2; finally a laser pulse on ion 1 reverses the first operation.

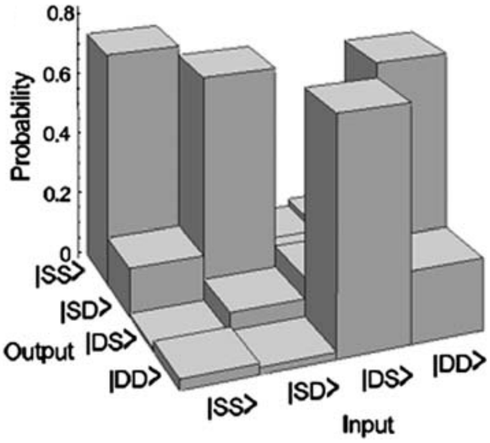

FIG. 10.6. Truth table for the Cirac-Zoller CNOT gate operation. $|g\rangle$ corresponds to the $S_{1/2}$ ground state and $|e\rangle$ to the $D_{5/2}$ metastable state of the $^{40}\text{Ca}^{+}$ ion. The fidelities (probabilities) represented are $|\langle\psi_{\text{experiment}}|\psi_{\text{ideal}}\rangle|^2$.

To implement a CNOT gate, it is possible to use a controlled-phase gate of amplitude π, which we shall note CPHASE(π), and apply single-qubit Hadamard gates, as can be seen in [**8**]. Let us first implement the CPHASE(π) gate[2].

[2] The CPHASE(δ) gate applies a phase shift to the target qubit only when the control qubit is in the state $|1\rangle$. Thus, it is represented by the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\delta} \end{pmatrix}.$$

The particular case $\delta = \pi$ is known as CMINUS [**7, 48**].

One can see [8] that, after applying a $\Omega_{\text{eff}}t = k\pi$ pulse, the Hamiltonian $H_{n,q}$ generates the following unitary transformation U on the ion-phonon system:

$$\begin{aligned}
U|g\rangle_n|0\rangle &= |g\rangle_n|0\rangle,\\
U|g\rangle_n|1\rangle &= \cos\left(\frac{k\pi}{2}\right)|g\rangle_n|1\rangle - i\sin\left(\frac{k\pi}{2}\right)|e_q\rangle_n|0\rangle,\\
U|e_q\rangle_n|0\rangle &= \cos\left(\frac{k\pi}{2}\right)|e_q\rangle_n|0\rangle - i\sin\left(\frac{k\pi}{2}\right)|g\rangle_n|1\rangle.
\end{aligned}$$

With the process

(i) π-pulse ($k = 1$) on the m-th ion on the 0-transition, $U_m^{k=1,q=0}$,
(ii) 2π-pulse ($k = 2$) on the n-th ion on the 1-transition, $U_n^{k=2,q=1}$,
(iii) π-pulse ($k = 1$) on the m-th ion on the 0-transition, $U_m^{k=1,q=0}$,

one obtains the transformation $U_{m,n}$, which is the CMINUS gate using as control qubit ion m and target qubit ion n:

$$U_{m,n} = U_m^{1,0}U_n^{2,1}U_m^{1,0}.$$

Indeed, it is easy to check the mapping on the ion$_m$-ion$_n$-phonon system:

$$\begin{array}{lclclcl}
|g\rangle_m|g\rangle_n|0\rangle & \to & |g\rangle_m|g\rangle_n|0\rangle & \to & |g\rangle_m|g\rangle_n|0\rangle & \to & |g\rangle_m|g\rangle_n|0\rangle,\\
|g\rangle_m|e_0\rangle_n|0\rangle & \to & |g\rangle_m|e_0\rangle_n|0\rangle & \to & |g\rangle_m|e_0\rangle_n|0\rangle & \to & |g\rangle_m|e_0\rangle_n|0\rangle,\\
|e_0\rangle_m|g\rangle_n|0\rangle & \to & -i|g\rangle_m|g\rangle_n|1\rangle & \to & i|g\rangle_m|g\rangle_n|1\rangle & \to & |e_0\rangle_m|g\rangle_n|0\rangle,\\
|e_0\rangle_m|e_0\rangle_n|0\rangle & \to & -i|g\rangle_m|e_0\rangle_n|1\rangle & \to & -i|g\rangle_m|e_0\rangle_n|1\rangle & \to & -|e_0\rangle_m|e_0\rangle_n|0\rangle.
\end{array}$$

The result is (like in Section 10.3.1) separable from the auxiliary phonon system and it can be seen that it really implements a CMINUS gate on the selected ions m and n from the trap.

The CNOT gate is obtained by applying a Hadamard gate both *a priori* and *a posteriori* of the CMINUS transformation on the target qubit [7].

Let us introduce the single ion transformations V_n and $V_n^\dagger$, which are represented in the basis $\{|g\rangle_n, |e_0\rangle_n\}$ as

$$V_n = \frac{1}{2}\begin{pmatrix} 1 & 1\\ -1 & 1 \end{pmatrix} = (V_n^\dagger)^\dagger.$$

Let us also observe that, by linearity,

$$\begin{aligned}
U_{m,n}|g\rangle_m(|g\rangle_n \pm |e_0\rangle_n)|0\rangle &= |g\rangle_m(|g\rangle_n \pm |e_0\rangle_n)|0\rangle,\\
U_{m,n}|e_0\rangle_m(|g\rangle_n \pm |e_0\rangle_n)|0\rangle &= |e_0\rangle_m(|g\rangle_n \mp |e_0\rangle_n)|0\rangle.
\end{aligned}$$

Finally, combining the two operations, we arrive at our goal:

$$\text{CNOT} = V_n^\dagger U_{m,n} V_n.$$

Indeed, a simple calculation proves the equivalence:

$$\begin{aligned}
|g\rangle_m|g\rangle_n|0\rangle &\to |g\rangle_m(|g\rangle_n - |e_0\rangle_n)|0\rangle \to |g\rangle_m(|g\rangle_n - |e_0\rangle_n)|0\rangle\\
&\to |g\rangle_m|g\rangle_n|0\rangle,
\end{aligned}$$

$$\begin{aligned}
|g\rangle_m|e_0\rangle_n|0\rangle &\to |g\rangle_m(|g\rangle_n + |e_0\rangle_n)|0\rangle \to |g\rangle_m(|g\rangle_n + |e_0\rangle_n)|0\rangle\\
&\to |g\rangle_m|e_0\rangle_n|0\rangle,
\end{aligned}$$

$$\begin{aligned}
|e_0\rangle_m|g\rangle_n|0\rangle &\rightarrow |e_0\rangle_m(|g\rangle_n - |e_0\rangle_n)|0\rangle \rightarrow |e_0\rangle_m(|g\rangle_n + |e_0\rangle_n)|0\rangle \\
&\rightarrow |e_0\rangle_m|e_0\rangle_n|0\rangle, \\
|e_0\rangle_m|e_0\rangle_n|0\rangle &\rightarrow |e_0\rangle_m(|g\rangle_n + |e_0\rangle_n)|0\rangle \rightarrow |e_0\rangle_m(|g\rangle_n - |e_0\rangle_n)|0\rangle \\
&\rightarrow |e_0\rangle_m|g\rangle_n|0\rangle,
\end{aligned}$$

where we also point out that the phonon state remains unchanged after the computation, in all cases, like the atom-atom entanglement described in Section 10.3.1.

With this we showed how a CNOT gate can be successfully made. This kind of gate is primordial because it creates entanglement between two qubits and it is also one of the pillars of quantum computation, since it forms, together with single qubit operations (as stated in Section 6.2.1), a set of universal gates. This implies that with such set of gates any quantum logic operation or quantum algorithm can be performed.

Conclusions

In this project we have presented the most important aspects of quantum information processing. We have interspersed with many remarks; aimed to be insightful commentaries, in order to orient ourselves in a world still filled with fundamental questions. Questions that arise from the so counter-intuitive reality quantum mechanics has proved to be, and that very few, if any, completely grasp.

Although some of these questions have already been answered, we still do not know how to operatively characterize entanglement, or if one-way functions can be built for a quantum computer, or if RSA can be generalized to a quantum computer, or which is the maximum error rate that allows secure quantum key distribution through a quantum channel, to mention some of them.

In the technological aspect, we have seen that even the implementation of a simple quantum operation poses a major challenge for its realization. Indeed, still very few qubits can be manipulated, in laboratories, and this is achieved with ‘computers´ that occupy a whole room. In fact, during a quantum computation, many things can go wrong. We have to find a trade-off between the isolation of the qubits from the environment and the ability to address them individually.

We have seen that there are, however, error correction mechanisms that help perform these operations more successfully. Nevertheless, they need the introduction of ancillary qubits, thus making the system bigger and more prone to decoherence. Thus, the error correction mechanism may need to be supervised by another error correction scheme, and so on. We have also proved that this procedure actually converges and quantum computation is ultimately possible if certain conditions are met.

These conditions, such as the threshold error probability, are tried to be fulfilled with several technologies, the most important of which we have discussed.

Despite the big leap between theory and practice, the real possibility of performing, some day, a quantum computation with a considerable number of qubits is probably what attracts most interest in the field that would be, otherwise, nothing but a mathematical curiosity.

We have also seen that quantum entanglement is a valuable physical resource, as it enables many novelties, and it is in the heart of quantum communication, quantum computation, quantum coding or quantum information in general. The non-classical correlations entangled pairs possess enable us to certify a sequence is truly random, to teleport a quantum state, to perform secure communication protocols, etc.

Quantum phenomena are also beginning to appear in other areas in which the approach has been typically classical. An example of this is microprocessors electronics: Although Moore’s law has been fairly accurate since it was coined in the 1970s, the constantly shrinking size of processors poses an impasse to the power of a classical computer and quantum phenomena are being more and more important, and need to be taken into account.

We have seen that quantum algorithms exist which outperform the best known classical algorithm. The example per excellence is Shor's Algorithm, which ultimately poses a threat to most of the communication protocols we are using nowadays if a quantum computer capable of manipulating a large enough number of qubits were to be built. Although it is not clear yet if quantum computation is more powerful than classical computation, we have seen that it is at least as powerful.

Quantum mechanics has also revolutionized the field of cryptography. Whereas classical cryptography is based on the assumption that the spy cannot solve a problem thought to be computationally hard, quantum cryptography assumes that the spy cannot break the laws of quantum mechanics. Another fundamental difference is that in classical cryptography there is no way to (directly) know if someone is listening and deciphering the communication; in quantum cryptography, the gain of information an eavesdropper could make introduces errors in the communication and his presence can be detected. Again, there is a big step from theory to practice, as we have seen that in a real scenario the spy has more tools at his disposal.

Future lines of development

During the development of this Thesis, many questions have been encountered.

In the context of quantum computing, more quantum algorithms appear every day. Which other problems can be more efficiently solved with a quantum computer? Is there some other scheme of quantum computation (such as quantum walks or adiabatic quantum algorithms, or measure-based quantum computation using cluster states, which are emerging tends) that can be exploited to lead to new results? In the complexity classification of problems for a quantum computer, the number of classes is broader, thus in a quantum context, more questions such as P=NP? can be asked.

In the characterization of entanglement, is there a necessary and sufficient condition that allows to decide it in bigger dimensions? Can entanglement witnesses be classified with respect to its optimality, at least in the decomposable case? How does one treat the general case?

With regard to quantum key distribution, how does one solve the problem of authentication? In the more realistic scenario, how can one be sure that all physical eavesdropping strategies are being considered and counteracted? Which is the maximum key rate that can be extracted for a given protocol and a given channel with its error probability?

In quantum coding, we have seen that CSS codes allow us to use classical codes to create a corresponding quantum code. This can be applied to Hamming codes or Reed-Solomon codes, for example. However, we have seen that degenerate codes exploit quantum features of the channel and thus they suggest that more efficient encoding could be performed.

Also, if degenerate codes improve the capacity of a quantum channel in the sense we described, as they provide one-way error correction, if we used them for two-way communication protocols, could we obtain higher key rates? Or could secure QKD still be possible with a higher noise level with the use of these codes?

About the technological aspects, the future lines of development are still broader: from the improvement of the photomultiplier detection efficiency to the perfect confinement of a ion trap. The first classical computer occupied a whole building; today we put one in our pocket, several orders of magnitude more efficient, faster and smaller. Today, a basic calculator can still outperform a quantum computer, which needs a whole laboratory and a considerable budget. In the future, we shall see...

References

[1] Acín, Antonio – Bae, Joonwoo: *Key distillation from quantum channels using two-way communication protocols*. Phys. Rev. A 75, 012334. 2007.

[2] Acín, Antonio: *Quantum communication*. Preprint. Seminar on Quantum Processing. 25-Feb-2010.

[3] Alicki, R.: *Field-Theoretical Methods*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 151-174.

[4] Augusiak, Remiguiusz – Tura, Jordi – Lewenstein, Maciej: *A note on the optimality of decomposable entanglement witnesses and completely entangled subspaces*. J. Phys. A: Math. Theor. 44, 212001. Fast Track Communication, IOP Publishing, 27-Apr-2011, 12pp.

[5] Benatti, Fabio – Fannes, Mark – Floreanini, Roberto – Petritis, Dimitri: (Eds.) *Quantum Information, Computation and Cryptography*. Lecture Notes in Physics, 808. Springer-Verlag, 2010.

[6] Benatti, Fabio: *Bipartite Quantum Entanglement*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 109-149.

[7] Benenti, Giuliano – Casati, Giulio – Strini, Giuliano: *Principles of Quantum Computation and Information, Volume 1: Basic concepts*. World Scientific, 2004.

[8] Benenti, Giuliano – Casati, Giulio – Strini, Giuliano: *Principles of Quantum Computation and Information, Volume 2: Basic tools and Special Topics*. World Scientific, 2007.

[9] Bennett, Charles H. – Brassard, Gilles: *Quantum Cryptography: Public key distribution and coin tossing*. Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India. pp. 175-179. Dec-1984.

[10] Blatt, Rainer – Häffner, H. – Roos, C. F. – Becher, C. – Schmidt-Kaler, F.: *Ion trap computing with Ca^+ ions*. Quantum Information Processing, 3. pp 61-73. 2004.

[11] Blatt, Rainer – Wineland, David: *Entangled states of atomic trapped ions*. Nature, 453, pp. 1008-1015. Jun-2008.

[12] Bruß, Dagmar: *Optimal Eavesdropping in Quantum Cryptography with Six States*. Phys. Rev. Lett. 81, 14. pp. 3018-3021. 1998.

[13] Bruß, Dagmar – Cirac, Juan Ignacio – Horodecki, Paweł– Hulpke, Florian – Kraus, Barbara – Lewenstein, Maciej – Sanpera, Anna: *Reflections upon separability and distillability*. Journal of Modern Optics, 49: 8. 2002. 1399-1418.

[14] Bruß, Dagmar – Meyer, T.: *Quantum Cryptography*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 277-308.

[15] Cirac, Juan Ignacio – Zoller, Peter: *Quantum computations with cold trapped ions*. Phys. Rev. Lett. 74, 20. pp 4091-4094. 1995.

[16] Clauser, John F. – Horne, Michael A. – Shimony, Abner – Holt, Richard A.: *Proposed experiment to test local hidden-variable theories*. Phys. Rev. Lett. 23, 15, pp. 880-884. 1969.

[17] Datta, Nilanjana: *Quantum Entropy and Information*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 175-214.

[18] Deutsch, David – Jozsa, Richard.: *Rapid solution of problems by quantum computation*, Proceedings of the Royal Society. London. Series A, Mathematical, Physical and Engineering Sciences, Volume 439 (1992), v+553 p.

[19] Diffie, Whitfield – Hellman, Martin E.: *New Directions in Cryptography*. IEEE Trans. Inf. Theory, 24. pp 339-351. 1976.

[20] DiVicenzo, David P. – Shor, Peter W. – Smolin, John A.: *Quantum channel capacity of very noisy channels*. Phys. Rev. A. 57:830. Feb-1998.

[21] Dyson, F. J.: *Advanced Quantum Mechanics*. Lecture notes of Relativistic Quantum Mechanics at Cornell University. Fall-1951.

[22] Ekert, Artur K.: *Quantum Cryptograhpy Based on Bell's Theorem*. Phys. Rev. Lett. 67, 6. pp. 661-663. 1991.

[23] Einstein, A. – Podolsky, B. – Rosen, N.: *Can quantum-mechanical description of physical reality be considered complete?*. Phys. Rev. **47** (1935), 777.

[24] Feynman, Richard P.: *Simulating physics with computers*. International Journal of Theoretical Physics, 21-6/7 (1982), 467-488.

[25] Feynman, Richard P.: *The Feynman Lectures on Computation* (edited by A. J. G. Hey and R. W. Allen). Addison-Wesley, 1996. Spanish translation: *Conferencies sobre computación*, Drakontos clásicos, Crítica, 2003.

[26] Genovese, M.: *Photonic Realization of Quantum Information Protocols*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 215-252.

[27] Hagley, E. – Maître, X. – Nogues, G. *et al.*: *Generation of Einstein-Podolsky-Rosen Pairs of Atoms*. Phys. Rev. Lett. 79, 1. 1997.

[28] Halmos, Paul: *Measure Theory*. Van Nostrand and Co. 1950.

[29] Home, Jonathan P. – Hanneke, David – Jost, John D. – Amini, Jason M. – Leibfried,Dietrich – Wineland, David J.: *Complete Methods Set for Scalable Ion Trap Quantum Information Processing*. Science, Vol 325. pp 1227-1230. 4-Sep-2009.

[30] Horodecki, Ryszard – Horodecki, Paweł– Horodecki, Michał– Horodecki, Karol: *Quantum Entanglement*. eprint quant-ph/0702225, 20-Apr-2007.

[31] Jaeger, Gregg: *Quantum Information –An overview*. Springer, 2007. xviii+284 p.

[32] Kempe, Julia – Vidick, T.: *Quantum Algorithms*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 309-342.

[33] Kern, Oliver – Renes, Joseph M.: *Improved one-way rates for BB84 and 6-state protocols*. eprint quant-ph/0712.1494v2, 5-Mar-2008.

[34] Kitaev A. Yu.: *Quantum measurements and the Abelian stabilizer problem*. Electr. Coll. Comput. Complex., 3rd Volume, 1995. Article no. 3, 22 pp.

[35] Kitaev, A. Yu. – Shen, A. H. – Vyalyi, M. N.: *Classical and quantum computation*. Graduate Studies in Mathematics, 47. American Mathematical Society, 2002. xiii+257 p.

[36] Kollmitzer, Christian – Pivk, Mario (Eds.): *Applied Quantum Cryptography*. Lecture Notes in Physics, 797. Springer-Verlag, 2010.

[37] Leinaas, Jon Magne – Myrheim, Jan – Sollid, Per ∅yvind: *Low-rank extremal positive-partial-transpose states and unextendible product bases*. Phys. Rev. A. 81, 062330. 2010.

[38] Lewenstein, Maciej – Kraus, Barbara – Cirac, Juan Ignacio – Horodecki, Paweł: *Optimization of entanglement witnesses*. Phys. Rev. A, Volume 62, 052310. The American Physical Society, 2000. 16 pp.

[39] Lewenstein, Maciej – Bruß, Dagmar – Cirac, Juan Ignacio – Kraus, Barbara – Kus, Marek – Samsonowicz, J. – Sanpera, Anna – Tarrach, R.: *Separability and distillability in composite quantum systems-a primer*. Journal of Modern Optics, 47: 14. 2000. 2481-2499.

[40] Maasen, Hans: *Quantum Probability and Quantum Information Theory*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 65-108.

[41] Maggiore, Michele: *A modern introduction to quantum field theory*. Oxford Master Series in Statistical, Computational, and Theoretical Physics. Oxford University Press, 2005

[42] McEliece, Robert J.: *The theory of information and coding*. Student Edition, Cambridge University Press. 2004.

[43] Melo, F. de – Buchleitner, A.: *Physical Realizations of Quantum Information*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 253-276.

[44] Mermin, David: *Quantum Computer Science: An Introduction*. Cambridge University Press, 2007. xv+220 p.

[45] Monz, Thomas *et al.*: *14-Qubit Entanglement: Creation and Coherence*. Phys. Rev. Lett. 106, 130506. Apr-2011.

[46] von Neumann, John: *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin. 1932. Translation into English by Beyer, Robert T.: *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, N.J. 1955.

[47] Mackey, George W.: *Mathematical Foundations of Quantum Mechanics*. Dover Publications. 2004. First Ed.: Benjamin. 1963.

[48] Nielsen, Michael A. – Chuang, Isaac L.: *Quantum Computation and Quantum Information*. Cambridge University Press, 2000 (5th printing 2005).

[49] Olmschenk, S. – Matsukevich, D. N. – Maunz, P. – Hayes, D. – Duan, L.-M. – Monroe, C.: *Quantum teleportation between distant matter qubits*. Science, 323, 486. 2009.

[50] Parthasarathy, K. R.: *Lectures on Quantum Computation, Quantum Error Correcting Codes and Information Theory*. Narosa Publishing House, 2006 (for the Tata Institute of Fundamental Research, international distribution by AMS). ii+128 p.

[51] Petz, D.: *Hilbert Space Methods for Quantum Mechanics*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 1-31.

[52] Pironio, S. – Acín, Antonio – Massar, S. *et al.*: *Random numbers certified by Bell's theorem*. Nature, 464, pp. 1021-1024. 2010.

[53] Preskill, John: *Lecture notes on quantum information and computation*. Available at *http://theory.caltech.edu/people/preskill*. 2004.

[54] Proakis, J. G. – Salehi, M.: *Communication Systems in Engineering* (2nd edition) Prentice Hall, 2002. p.586.

[55] Rivest, R. L. – Shamir, A. – Adleman, L.: *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*. Commun, ACM, 21. pp 120-134. 1978.

[56] Rué, Juanjo – Xambó, Sebastià: *Mathematical essentials of quantum computing*. Preprint. Seminar on Quantum Processing. 25-Feb-2010.

[57] Rukhin, Andrew *et al.*: *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. NIST Special Publication 800-22. Revision 1a. Apr-2010.

[58] Shannon, Claude Elwood – Weaver, Warren: *The Mathematical Theory of Communication*. The Universirty of Illinois Press. 1949.

[59] Shannon, Claude Elwood: *Communication Theory of Secrecy Systems*. Bell System Technical Journal, 28-4. pp 656-715. Oct-1949.

[60] Shor, Peter W.: *Algorithms for quantum computation. Discrete log and factoring*. Proceedings of the 35th IEEE Annual Symposium of the Foundations of Computer Science (FOCS). IEEE Computer Society, Los Alamitos. pp 124-134. 1994.

[61] Shor, Peter W. – Smolin, John A.: *Quantum error correcting codes need not completely reveal the error syndrome*. eprint quant-ph/9604006v2, Apr-1996.

[62] Shor, Peter W.: *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*. SIAM Journal on Computing, 26-5 (2005), 1484-1509.

[63] Smith, Graeme – Smolin, John A.: *Degenerate quantum codes for Pauli channels*. Physical Review Letters, 98:030501-4. 2007.

[64] Stephen, Jordan: *Quantum Algorithm Zoo*. Updated quantum algorithm list available at *http://math.nist.gov/quantum/zoo/*. 2011.

[65] Stolze, Joachim – Suter, Dieter: *Quantum Computing*. Wiley-VCH. 2008.

[66] Suhov, Y.: *Classical Information Theory*. Lecture Notes in Physics, 808. Springer-Verlag, 2010. 33-64.

[67] Verdú, Sergio – McLaughlin, Steven: *Information Theory: 50 Years of Discovery*. IEEE Press, 1999.

[68] Vernam, Gilbert S.: *Cipher Printing Telegraph Systems For Secret Wire and Radio Telegraphic Communications*. Journal of the IEEE, 55. pp 109-115. 1926.

[69] Weinberg, Steven: *The quantum theory of fields. Volume I: Foundations*. Cambridge University Press. 1995.

[70] Wick, G. C.: *The evaluation of the collision matrix*. Phys. Rev. 80, 268. 1950.

[71] Xambó, Sebastià: *Error-correcting codes*. Lecture notes available at *http://www-ma2.upc.es/sxd/sxdeng.html*. 2010.

List of Figures

Notation

In this section is listed the notation most generally used in this Thesis. Terms which are defined are underlined, whereas *terms* to be emphasized are written in italics.

a^*	:	Complex conjugate
$\widehat{a}$	:	Operator
$\boldsymbol{a}$	:	Vector
a	:	Matrix representation of $\widehat{a}$
a^T	:	Transposition of matrix a
$a^\dagger$	:	Hermitian transposition of matrix a
$\mathbb{I}_n$	:	Identity matrix or operator acting on a $n-$ dimensional space
$\mathbb{R}$	:	The field of real numbers
$\mathbb{C}$	:	The field of complex numbers
$\mathbb{Z}$	:	The ring of integer numbers
$\mathbb{N}$	:	The set of natural numbers
$\mathbb{Z}_m$	:	The ring $\mathbb{Z}/(m)\mathbb{Z}$
$\mathbf{P}$	:	A classical probability function $\mathbf{P} : \sigma \longrightarrow [0,1]$ in a classical probability space $(\Omega, \sigma, \mathbf{P})$
$\mathbb{E}$	:	Expectation of a random variable
$\mathfrak{S}_n$	:	The group of permutations of n elements
$\#A$	:	Cardinal of set A
$C^\perp$	:	Dual of code C
$d^n\boldsymbol{x}$	:	Measure over $\mathbb{R}^n$
$\delta_{i,j}$	:	Kronecker's delta function
$\Re(a)$	:	Real part of $a \in \mathbb{C}$
$\Im(a)$	:	Imaginary part of $a \in \mathbb{C}$

List of Acronyms

AEP	Asymptotic Equipartition Property
AES	Advanced Encryption Standard
BB84	Bennett & Brassard, 1984
BS	Beam-Splitter
C	Constant
CAR	Canonical Anti-Commutation Relation
CCD	Charge Coupled Device
CCR	Canonical Commutation Relation
CHSH	Clauser-Horne-Shimony-Holt
CMINUS	Controlled π-Phase
CNOT	Controlled NOT
CPHASE	Controlled Phase
CPT	Completely Positive, Trace preserving map
CQED	Cavity Quantum ElectroDynamics
CSS	Calderbank, Shor and Steane
CZ-CNOT	Cirac-Zoller CNOT
DEW	Decomposable EW
E	Exponential
EPR	Einstein-Podolsky-Rosen
EW	Entanglement Witness
GHZ	Greenberger, Horne and Zeilinger
HSP	Hidden Subgroup Problem
IID	Independent, Identically Distributed
I&R	Intercept & Resend
NPT	Non-positive under Partial Transposition
P	Polynomial
PMT	PhotoMultiplier Tube
PNS	Photon Number Splitting
POVM	Positive Operator-Valued Measure
PPT	Positive under Partial Transposition
QFS	Quantum Fourier Sampling
QFT	Quantum Fourier Transform
QKD	Quantum Key Distribution
RNG	Random Number Generator
RSA	Rivest, Shamir and Adleman
SNR	Signal to Noise Ratio
SP	Super Polynomial
UPB	Unextendible Product Basis