Exploring equilibration in fermionic systems:
A connection with minimum distance codes

Jose Alejandro Montaña Cortes

Universidad de los Andes
Faculty of Science, Department of Physics
Bogotá, Colombia
2020

Exploring equilibration in fermionic systems: A connection with minimum distance codes

Jose Alejandro Montaña Cortes

Supervisor:

Alonso Botero Mejía

A thesis submitted in fulfilment of the requirements for the degree of:

Master's Degree in Physics

In the research group of:

Quantum Mechanics and Information Theory

Universidad De Los Andes

Faculty of Science , Departament of Physics

Bogotá, Colombia

2020

To everyone that was there to listen to me.

Resumen

La tipicidad canónica ha surgido como una alternativa a los fundamentos de la mecánica estadística, logrando explicar desde una perspectiva de la información cuántica, fenómenos como el de la termalización, que emerge como consecuencia del entrelazamiento entre el sistema y su ambiente. Posteriores trabajos han mostrado que esta particular forma de abordar el problema proporciona una idea del mecanismo por el cual se alcanza el equilibrio en sistemas cuánticos. Dentro del marco de estas ideas, promedios temporales juegan un rol principal a la hora de entender cómo estados reducidos obtenidos a partir de un estado puro, alcanzan el equilibrio. Esto nos conduce a preguntarnos: ¿Es posible encontrar estados reducidos provenientes de un estado puro no estacionario, que automáticamente estén equilibrados, es decir, estados reducidos cuyo equilibrio se alcance de forma inmediata?, y en caso de ser así, ¿es posible determinar el tamaño del conjunto de estados cumpliendo esta propiedad?

En el presente trabajo, el uso de herramientas de teoría de código, específicamente, códigos aleatorios fermiónicos de distancia mínima, proporciona una forma diferente de entender los sistemas fermiónicos reducidos. Confirmando así, la existencia de estados reducidos que permanecen en equilibrio para estos sistemas. En este sentido, se explora el espacio de Hilbert generado por el conjunto de estos estados, concretamente, se caracteriza el espacio de Hilbert asociado a estos estados por medio de un exponente de error, el cual proporciona una ley de grandes desviaciones en el sistema. De esta forma, se prueba que el tamaño del espacio de Hilbert asociado a los estados, que cumplen la propiedad de equilibrarse de forma inmediata, es exponencialmente grande.

Palabras Clave: Tipicidad, Sistemas fermiónicos, códigos aleatorios fermiónicos de distancia mínima .

Abstract

Canonical typicality has emerged as an alternative to the foundations of statistical mechanics. It has been able to explain from a viewpoint of quantum information theory, how thermalisation emerge from entanglement between the system and its environment. Later results have shown that the ideas used in typicality, provide an explanation to the mechanism of the evolution towards equilibrium for large quantum systems. Within this framework of ideas, time averages plays a major rol when it comes to understand how reduced states, obtained from a pure states, can reach equilibrium. This led us to ask ourselves, is it possible to find reduced states obtained from a non stationary pure state, such that they equilibrate instantly, that is, reduced states such that immediately reach its equilibrium state? If so, is it possible to measure the size of the space fulfilling this property?

In the present work, tools of code theory, specifically fermionic minimum distance codes, provide an alternative understanding of fermionic systems. Specifically, we characterised the size of the Hilbert space generated by the states, such that fulfil this property, throughout an error exponent, which provides a characterisation of the large deviation law within the system. Hence, we prove that the Hilbert space of states fulfilling the property of reaching its equilibrium instantly is indeed exponentially large.

Key Words: Typicality, fermionic systems, fermionic minimum distance codes.

Agradecimientos

Este trabajo es el resultado de un trabajo de investigación que empezó por estudiar un sistema específico y terminó proporcionando una visión de algo mucho más general. Aunque este documento no plasme todo lo investigado a los largo de mi proyecto de la maestría, si creo que muestra el esfuerzo que esta investigación implicó. Por esto quisiera expresar mis agradecimientos a todas las personas con las que en algún momento crucé palabra y que me enseñaron mucho más de lo que la academia me enseñó. En especial quisiera agradecer al profesor Alonso Botero por su paciencia y dedicación para conmigo, pues a pesar de nuestras diferencia, esta investigación no hubiera sido posible de no haber sido por él.

Quiero agradecer especialmente a Valentina Cobo por haberme apoyado y aconsejado durante todo el proceso que tuve en mi maestría. Adicionalmente quisiera agradecer a Mario Henao, Mauricio Sevilla y Julian Triana por haber dedicado parte de su tiempo para escucharme y ayudarme.

A todos ustedes es a quien debo este trabajo.

Declaration of Authorship

I, Jose Alejandro Montaña Cortes, declare that this thesis titled,“ *Exploring equilibration in Fermionic systems: A connection with minimum distance codes*” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Content

Introduction

In the last two decades, a wave of works in quantum information theory has proven to be a new and objective form of understanding the foundations of statistical mechanics [1–7]. Particularly, alternative considerations about the foundations of statistical mechanics proposed by Posescu et. al. [1], have shown that reliability on subjective randomness [8], ensemble averaging [9] or time averaging [10] are not required to understand the emergence of thermalisation; instead, a quantum information perspective [11] provides an alternative answer to the foundations of statistical mechanics, from a purely quantum point of view, which does not rely on any ignorance probabilities in the description of the state. This precise way of tackling the problem is intimately related to empiricism, for example: Whenever you plan your day, or you go out for a walk, the last thing that goes through our mind is to get smacked by a meteor, we know it may happen, but we know that it is not something "normal" to occur. That is why it is "typical" to plan our lives without even bothering by getting hit by a meteor. Similarly, what Popescu et. al. [1] proved is that if we consider a quantum pure state, subject to a global constraint (e.g a constant energy), the "typical" thing to happen is that the reduced state of the system is very close to the canonical mixed state. That is, almost every reduced state obtained from a quantum pure state will approximately coincide with the canonical thermodynamic state [12, 13].

To provide a more precise argument, consider as the universe ($\mathcal{U}$) the system ($\mathcal{S}$) together with a sufficiently large environment ($\mathcal{E}$), in a quantum pure state. Due to entanglement between the system and the environment and properties of high dimensional Hilbert spaces, system thermalisation appears as a local generic property of pure states of the universe subjected to a global constraint. This result is known as the *general canonical principle*, or informally as *canonical typicality* and is considered to be an important result when understanding statistical mechanics in quantum systems. Specifically, what this principle tells us is that whenever we look at a sufficiently small system, compared to its environment, the reduced state of the system will approximately correspond to the thermal state [1, 2, 4, 14], therefore suggesting that thermalisation occurs as a generic local property of pure states of the universe.

It should be emphasised that the results in typicality apart from providing a general viewpoint of thermalisation, the nature of those arguments is kinematic, rather than dynamic. That is, the particular unitary evolution of the global state is never considered, and thermalisation is not proven to happen; instead, the key ingredient is Levy's Lemma [15, 16], which plays a similar role to the law of large numbers and governs the properties of typical states in large-dimensional Hilbert spaces [1], and thus provides a powerful tool to evaluate functions of randomly chosen quantum states. We stress here that these ideas were not only proposed by Popescu et. al. [1]; contemporaneously with them, Gemmer. et. al. [14], as well as Goldstein et. al. [2], proposed similar ideas, in which heuristic arguments are used to prove canonical typicality, and exhibit an explicit connection between reduced states an the micro-canonical density matrix at a suitable total energy E. However, the result obtained by Popescu et. al. establishes canonical typicality in a general way by invoking the Levy's Lemma [1, 15, 16], For that reason, the viewpoint we discuss here is mostly based on the one proposed by Popescu.

With the purpose of extending typicality beyond the kinematic viewpoint and address the dynamics of thermalisation, we enquire under what conditions the state of the universe will evolve into the typical region of its Hilbert Space in which its subsystems are thermalised and remain in that space for most of its evolution. Motivated by previous results heading this direction [3, 17] and the fact that, from typicality arguments is possible to show that the overwhelming majority of states in the universe bring the system to the canonical mixed state, Linden et. al. [18] explore whether thermalisation could happen as a universal property of quantum systems. Thus, by using arguments based on ideas of typicality, reaching equilibrium can be shown to be a typical property of large quantum systems. In this framework, dynamical aspects are addressed to explore the evolution that drives systems to equilibrate, moreover, to study under what circumstances systems reach equilibrium and how much they fluctuate about the equilibrium state. A series of results in [18–20] suggest that under mild conditions, any subsystem of a sufficiently large system will reach equilibrium and fluctuate around it at almost every time. The only conditions required are that the Hamiltonian has no degenerate energy gaps, and that the state of the universe contains sufficiently many energy eigenstates. These conditions are fulfilled for most physical situations, in fact all but a set of measure zero of Hamiltonians have non-degenerate energy gaps [18].

Even though thermalisation seems to be a very straightforward process, it is quite difficult to formalise an explanation to it. A closer look, reveals that thermalisation is composed of many different aspects that have to be inspected in detail, and where *equilibration*, *bath state independence*, *subsystem state independence* and *the Boltzmann form of the equilibrium state* play a prominent role [18]. First, *equilibration* is the process in which the system reaches a particular state and remains in that state or close to it. Whenever we refer to equilibration, note that any particular state is not inferred and in general it does not need to be a thermal state. Second, *bath state independence* refers to the fact that the equilibrium state of the system should not depend on the precise initial state of the Bath. That is, only macroscopic parameters are needed to describe the bath [18], for example, its temperature: In the moment equilibrium is reached, that state should only depend on the temperature of the bath. Third, *subsystem state independence* refers to the fact that the equilibrium state reached by the system, should be independent of its initial state. Finally, the *Boltzmann form of the equilibrium state* refers to the form of the equilibrium state $\rho_{\mathcal{S}} = \frac{1}{Z}\exp(-\frac{H_{\mathcal{S}}}{k_B T})$, which is known as a Boltzmanninan form of equilibrium. Note that equilibration is then a more general process which can depend on different parameters such as initial conditions in an arbitrary way, whereas thermalisation does not.

Realising that thermalisation is compound by the afore-mentioned elements, let us clarify an important distinction between thermalisation and equilibration, indeed, we will consider equilibration as a general quantum phenomenon that may occur in situations other than those associated with thermalisation. By using this decomposition of thermalisation, Linden et. al. were able to prove the first two elements mentioned above (Equilibration and bath state independence). Namely, they prove not only that reaching equilibrium is a universal property of quantum systems but that this equilibrium state does not depend on the precise details of the bath state, but rather on its macroscopic parameters [18].

Up to this point, we have shown that typicality has been proven to be an extremely useful alternative way of studying thermalisation in quantum systems and understanding the foundations of quantum statistical mechanics. Nonetheless, a closer look to typicality will derive in a property that we consider could yield to further insights to understand equilibration in quantum systems. To illustrate our ideas, consider two different orthogonal pure states, $|\psi_1\rangle$ and $|\psi_2\rangle$, living in the same Hilbert subspace $\mathcal{H}_R$ ($|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}_R$), the one that is obtained by imposing a global constraint, denoted by R, over the universe. From typicality we know that the reduced state of $|\psi_1\rangle$ and $|\psi_2\rangle$ approximately leads to the same state, that is,

$$\mathrm{Tr}_{\mathcal{E}} |\psi_1\rangle \langle\psi_1| \approx \mathrm{Tr}_{\mathcal{E}} |\psi_2\rangle \langle\psi_2| \approx \Omega_{\mathcal{S}}, \tag{1}$$

where $\Omega_{\mathcal{S}}$ corresponds to the canonical state of the system. Thus, we could consider a third state $|\psi_3\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle$ which is a generic linear combination of $|\psi_1\rangle$ and $|\psi_2\rangle$. We have then that its reduced state will also lead us to the canonical state, meaning that cross terms obtained in the reduced state associated with $|\psi_3\rangle$ should somehow vanish. Explicitly, when we compute the density matrix associated with the state $|\psi_3\rangle$ we have

$$|\psi_3\rangle \langle\psi_3| = |c_1|^2 |\psi_1\rangle \langle\psi_1| + |c_2|^2 |\psi_2\rangle \langle\psi_2| + c_1^* c_2 |\psi_2\rangle \langle\psi_1| + c_2^* c_1 |\psi_1\rangle \langle\psi_2| , \tag{2}$$

and hence its reduced state reads

$$\mathrm{Tr}_{\mathcal{E}} |\psi_3\rangle \langle\psi_3| \approx \Omega_{\mathcal{S}} \approx \Omega_{\mathcal{S}} + c_1^* c_2 \,\mathrm{Tr}_{\mathcal{E}} |\psi_2\rangle \langle\psi_1| + c_2^* c_1 \,\mathrm{Tr}_{\mathcal{E}} |\psi_1\rangle \langle\psi_2| , \tag{3}$$

where the condition of normalisation was used ($|c_1|^2 + |c_2|^2 = 1$). Notice that the cross terms in equation (3) should therefore approximately vanish in order to satisfy the relation. Namely, the condition which has to be satisfied in order to keep the equality is $\mathrm{Tr}_{\mathcal{E}} |\psi_2\rangle \langle\psi_1| = \mathrm{Tr}_{\mathcal{E}} |\psi_1\rangle \langle\psi_2| = 0$. Since this condition tell us that off-diagonal terms approximately vanish this brought to our mind some previous ideas proposed by Srednicki et. al. [21–23] in eigenstate thermalisation hypothesis (ETH), where the off-diagonal terms are expected to be stochastic quantities with mean zero and an amplitude that is exponentially small on the number of degrees of freedom of the system. That is, the off-diagonal terms are expected to be near zero for a system with a large number of degrees of freedom. For our case, we have something similar but instead of having that the off-diagonal terms are near to zero, we have that after taking the partial trace, those terms become approximately zero. Although we are aware that the condition $\mathrm{Tr}_{\mathcal{E}} |\psi_2\rangle \langle\psi_1| \approx \mathrm{Tr}_{\mathcal{E}} |\psi_1\rangle \langle\psi_2| \approx 0$ might be related with ETH in some way, due to the impossibility of making an explicit connection between them, we name it *ultra-orthogonality*,

$$\mathrm{Tr}_{\mathcal{E}} |\psi_i\rangle \langle\psi_j| \approx 0, \qquad i \neq j. \tag{4}$$

This particular name was given since when we compute the partial trace over the exterior product of $|\psi_i\rangle$ and $|\psi_j\rangle$ it becomes zero, thus we consider this name provides the idea of having orthogonality over partial traces.

Ultra-orthogonality, can be shown to be related with the equilibration of systems. Consider a time dependent state $|\Psi(t)\rangle$, we can expand this state onto its energy eigenstates as

$$|\Psi(t)\rangle = \sum_k c_k e^{-iE_k t} |E_k\rangle\,, \tag{5}$$

where $\sum_k |c_k|^2 = 1$; hence,

$$\rho(t) = \sum_{k,\ell} c_k c_\ell^* e^{-i(E_k - E_\ell)t} |E_k\rangle \langle E_\ell| = \underbrace{\sum_k |c_k|^2 |E_k\rangle \langle E_k|}_{\omega} + \sum_{k\neq\ell} c_k c_\ell^* e^{-i(E_k - E_\ell)t} |E_k\rangle \langle E_\ell|\,, \tag{6}$$

where the fist term ω is time-independent, so when we look at the reduced state of $\rho(t)$, defined as $\rho_{\mathcal{S}}(t) = \mathrm{Tr}_{\mathcal{E}} |\Psi(t)\rangle \langle \Psi(t)|$,

$$\rho_{\mathcal{S}}(t) = \omega_{\mathcal{S}} + \sum_{k\neq\ell} c_k c_\ell^* e^{-i(E_k - E_\ell)t}\, \mathrm{Tr}_{\mathcal{E}} |E_k\rangle \langle E_\ell|\,, \tag{7}$$

where $\omega_{\mathcal{S}} = \mathrm{Tr}_{\mathcal{E}}\, \omega$. Thus, if all states in equation (7) satisfy the ultra-orthogonality property, then the reduced state would automatically be stationary, meaning that it will be time independent.

Remembering that equilibration is a general process in which the state remains for the almost every time (time independent), allows us to see that if the states in equation (7) satisfy ultra-orthogonality, then ultra-orthogonality will be related with equilibration as an immediate phenomenon. Moreover, from typicality we know that the overwhelming majority of states are such that its reduced state approximately coincide with the thermal state, thus, one can anticipate that for most cases the equilibrium state that we will be studying will be the canonical thermal state. It would therefore be interesting to fully study the property of ultra-orthogonality to provide a better comprehension of what is behind this phenomenon and provide what may appear to be an alternative instantaneous mechanism for equilibration. However, in this work we will not address the general problem of studying ultra-orthogonality in any quantum system; instead, we will be discussing a particular case of ultra-orthogonality. Here we will describe the reduced state of a pure and fully dynamical state of the universe, such that its reduced states, sufficiently small compared its environment, are automatically stationary. More specifically, we will be discussing the case when the cross terms in (7) are *exactly* equal to zero ($\mathrm{Tr}_{\mathcal{E}} |E_k\rangle \langle E_\ell| = 0$) in the special case of systems that their Hamiltonians represent interacting systems that can be mapped, under appropriate approximations (or transformations), to Hamiltonians which are quadratic in fermionic operators of the form

$$\hat{H} = \sum_{ij} C_{ij} \hat{a}_i^\dagger \hat{a}_j + \sum_{ij} \left(A_{ij} \hat{a}_i^\dagger \hat{a}_j^\dagger + \mathrm{h.c.} \right), \tag{8}$$

where i, j run from 1 to N, the number of modes in the system.

One of the first questions one might wonder is, "are there few or conversely, many states that fulfil the property that the cross terms in (7) are exactly equal to zero?" and consequently "what is the size of the Hilbert subspace associated with these states?". We will approach these questions by using some ideas taken from random minimum distance codes adapted to fermionic systems, and we will show through a concept taken from code theory, the *random-coding exponent*, that there are exponentially large Hilbert subspaces in which all its states fulfil the condition $\mathrm{Tr}_{\mathcal{E}} |E_k\rangle \langle E_\ell| = 0$.

An insight to clarify how techniques from code theory can be used to study ultra-orthogonality starts by noting that fermionic states can be interpreted as occupation numbers n_i in the different modes, where $n_i \in \{0, 1\}$. Namely, we will be looking at systems of N modes and we will call the subsystem the set of L modes, where $N \gg L$.For this case, we have that the state of the system will be represented as a binary sequence $|\vec{n}\rangle = |n_1, n_2, \ldots\rangle$ in the Fock space. Now consider the states $|\vec{n}_1\rangle$ and $|\vec{n}_2\rangle$ that are mutually orthogonal eigenstates of the Hamiltonian (8). We denote by $\vec{n}_1$ ($\vec{n}_2$) the sequence of excitations present in the correspondent state $|\vec{n}_1\rangle$ ($|\vec{n}_1\rangle$). Now we consider the error vector $\vec{e}_{12} = \vec{n}_1 + \vec{n}_2$, which its entries are zero whenever the $i-$th element of both sequences $(\vec{n}_1, \vec{n}_2)$ is equal and one otherwise, observe that the higher the energy, the more excitations we have in the sequences, so the vector $\vec{n}_1 + \vec{n}_2$ will be very likely to have more ones when we increase the energy. Denoting by d the number of ones in the vector $\vec{e}_{12}$, as we will explain in detail in this document, it is possible to prove that when $d > L$,

$$\hat{X}_{12} = \mathrm{Tr}_{N-L} |\vec{n}_1\rangle \langle \vec{n}_2| = 0, \tag{9}$$

where Tr_{N-L} represents the partial trace over the $N - L$ elements. This condition of getting crossed terms equals to zero at certain distance d brought to our minds the idea of minimum distance codes, codes that have the property of perfectly correct errors when the number of errors do not exceed $\lfloor (d-1)/2 \rfloor$. Hence, our idea is to look at the minimum distance code $\mathfrak{C} = \{x^{(1)}, x^{(2)}, \ldots, x^{(2^k)}\}$ with 2^k *codewords*, such that their respective states $\{|x^{(1)}\rangle, |x^{(2)}\rangle, \ldots, |x^{(2^k)}\rangle\}$ are mutually ultra-orthogonal. We will be concern with the size of the largest minimum distance code that can be built to alternatively answer what is the size of the largest Hilbert subspace $\mathcal{H}_{\mathfrak{C}}$ spanned by these vectors ($\mathcal{H}_{\mathfrak{C}} = \mathrm{Span}\left(\{|x^{(1)}\rangle, |x^{(2)}\rangle, \ldots, |x^{(2^k)}\rangle\}\right)$). That is, if we look at a state $|\psi\rangle \in \mathcal{H}_{\mathfrak{C}}$, it can be expanded in terms of the basis vectors $|x^{(i)}\rangle$

$$|\psi\rangle = \sum_{\vec{n}\in\mathfrak{C}} \psi(\vec{n}) \, |\vec{n}\rangle \,, \tag{10}$$

and then the reduced state of the system will be

$$\rho_{\mathcal{S}}(\psi) = \mathrm{Tr}_{\mathcal{E}}(|\psi\rangle \langle\psi|) = \sum_{\vec{n}\in\mathfrak{C}} |\psi(\vec{n})|^2 \, \mathrm{Tr}_{\mathcal{E}} \, |\vec{n}\rangle \langle\vec{n}| \,, \tag{11}$$

which means that the reduced state is immediately stationary. Therefore if we are able to find the largest minimum distance code, we will be alternatively answering what is the largest size of the Hilbert subspace spanned by the codewords $\{x^{(i)}\}$. Fortunately, previous works [24] provide the estimate size of the largest random minimum code that can be built in a memoryless binary symmetric channel, calculations that can be adapted to our purpose and that will let us show that is possible to build an exponentially large Hilbert subspace in which ultra-orthogonality holds exactly.

This document is divided in four chapters: In the first chapter, we will introduce canonical typicality and its relation to statistical mechanics. Starting with a discussion of the thermodynamic entropy and the main differences between Boltzmann's and Gibb's approach to thermodynamics. Next we introduce typicality and its foundations, and finally a detailed explanation of the problem of ultra-orthogonality. In the second chapter, we will provide all the theoretical background concerning fermionic systems and code theory, in particular, we will be discussing the case of the one dimensional XY model, and we will finish with the introduction of some important results in code theory. In the third chapter, we present our main results which are an explicit connection between a special kind of fermionic systems and random minimum distance codes. Also, we will compute an estimate size of the Hilbert subspace associated with the states that fulfil ultra-orthogonality. Finally, the fourth chapter contains the conclusions and further perspectives of our work.

Chapter 1

Canonical typicality and its connection to thermodynamics

Fundamental questions concerning the foundations of quantum statistical mechanics have been discussed and remain a debatable subject [13]. In these questions the role of probabilities, entropy, the relevance of time averages and ensemble averages to individual physical systems are discussed thoroughly [14] to answer whether or not they are needed to formalise statistical mechanics. One of the most controversial issues is the validity of the postulate of equal a priori probability, postulate which can not be proven [13] and has been used since 1902 by Gibbs [25], who introduces the emergence of lack of knowledge to formalise the ideas of classical statistical mechanics. Along this chapter, we will discuss some of the ideas based on typicality addressed by several authors [2, 4, 14], who have abandoned the unprovable aforementioned postulate and have replaced it with a new viewpoint, which is purely quantum, and which does not rely on any ignorance probabilities in the description of the state.

This chapter is divided in three parts: First, we are going to discuss the thermodynamic entropy. To do this, we are going to compare the Gibbs entropy as well as the Boltzmann entropy, and we are going to argue that the correct thermodynamic entropy corresponds to the Boltzmann entropy. Also the quantum extension of these ideas are going to be discussed for the case of the quantum Gibbs entropy. The second part of the chapter, will be dedicated to understand the role of entanglement in connection with Statistical mechanics. We introduce the concentration phenomenon of canonical typicality, from which thermalisation will emerge as a consequence of typicality [1, 4]. Finally, we will study a dynamical point of view of thermalisation proposed by Linden et. al. [18], to show that equilibration emerge as a generic property of local quantum systems. This will lead us to the main subject of our work, the property we named ultra-orthogonality. We will look under what conditions ultra-orthogonality is expected to happen and we will explicitly mention the problem that we will tackle in this document.

1.1. The idea of thermodynamic entropy

Science is often presented as a collection of universally accepted knowledge and discoveries, in which disagreement among scientists is often downplayed. Specifically, in physics many questions have been a matter of disagreement, where not only questions related with new discoveries have played a main role in the discussions, but also questions concerning concepts that have been taken for granted in the books, such as the interpretation of the quantum mechanics. Particularly, when we look at the foundations of statistical mechanics, two different formulations of entropy are often presented in the literature. The first one, proposed by Boltzmann [26], which provides a definition of thermodynamic entropy for an individual system, and the other one, proposed by Gibbs [25], which gives an entropy definition of a probability distribution over the phase space.

It is stated that the Gibbs entropy gives the correct thermodynamic entropy [27], since it yields to the correct thermodynamic predictions, while Boltzmann $H-$theorem is correct only in the case of ideal gases. However, there is a school of thought which holds that Boltzmann expression is directly related to the entropy, and the Gibbs' one is simply erroneous and misleading [28].

Is not new that statistical physics based on the Gibbs interpretation has provided high accurate results that has yielded to the correct thermodynamic predictions. Hence, the problem behind the discussions relating the foundations of statistical mechanics are not referring to its usefulness, but instead, with subtle differences in the the interpretations behind these theories.

1.1.1. Boltzmann entropy

Let $X = (\vec{q}_1, \ldots, \vec{q}_N; \vec{p}_1, \ldots, \vec{p}_N)$ be the *microstate* of a classical system at a time t, consisting of a large number N of identical particles forming a gas in a box Λ. The evolution of the system is then determined via Hamilton's equations of motion

$$\frac{\mathrm{d}\vec{q}_i}{\mathrm{d}t} = \frac{\partial H}{\partial \vec{p}_i}, \qquad \frac{\mathrm{d}\vec{p}_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \vec{q}_i}, \tag{1.1}$$

with $H(\vec{q}_1, \ldots, \vec{q}_N; \vec{p}_1, \ldots, \vec{p}_N)$ the Hamiltonian function. Since the energy E is a constant, the evolution of the system is then confined to a set of Hamiltonians Ω_E such that fulfil the restriction $\Omega_E = \{(\vec{q}, \vec{p}) \in \Omega(\Lambda) | H(\vec{q}, \vec{p}) = E\}$, where we represent by $\Omega(\Lambda)$ the set of all possible states in the phase space. Now consider the subset Γ of Ω_E ($\Gamma \subset \Omega_E$), as the set of all phase points that "look macroscopically similar" to X. In other words, every phase point X has an associated *macrostate* $\Gamma(X)$ consisting of phase points that are macroscopically similar to X.

Thus, if we reticulate the one-particle phase space ($\vec{q}, \vec{p}$-space) into macroscopically small but microscopically large cells Δ_α, over each cell we can specify the number n_α of particles on each cell. By doing so, we will end up looking at a histogram which is a deterministic function of the microstate of the system and that specifies a macrostate over time. Notice that the histogram we are building is not a probability distribution, in fact, it corresponds to an instantaneous occupation number which tells us how many particles are in certain state, hence, probabilities are not needed to interpret the macrostate. However, the value of the histogram remains unknown for us, because we have no clue what the initial conditions were, hence, this is the moment probability plays a crucial role in Boltzmann's description.

One of the Boltzmann's great achievement in [26], was to arrive at an understanding of the meaning of the *Boltzmann entropy* as a measure of the set of all phase points that look macroscopically like X ($\Gamma(X)$). Explicitly he found that the quantity[1]

$$S(X) = k_B \log |\Gamma(X)|, \tag{1.2}$$

gives an expression for the thermodynamic entropy, to which the Second law refers. Where k_B is the Boltzmann's constant, and $|.|$ denotes the volume given by the Lesbege measure onto Ω_E.

In his paper [26] Boltzmann asks himself about the most common histogram that appears for a given macrostate. What he proved is that the vast majority of the points in the phase space have the property to end up looking as the histogram found when maximizing S, that

[1] Boltzmann used the notation H to refer to the entropy, we do not use it to not confuse it with the Hamiltonian.

is the histogram that corresponds to the equilibrium macrostate. Thus, Ω_E will consist almost entirely of phase points in the equilibrium macrostate Γ_{Eq}, with few exceptions, whose totality has volume of the order 10^N relative to that of Ω_E. Moreover, for non-equilibrium phase points X of energy E, the Hamiltonian dynamics, governing the motion of X_t arising from X, would have to be exceptionally special to avoid reasonably quick carrying X_t to Γ_{Eq} and keep it there for a long time.

Even though Boltzmann's ideas provide the correct definition of thermodynamic entropy, Boltzmann entropy fails when we consider interacting systems [29, 30]. The work of Boltzmann in the next years [31, 32] was dedicated to include the interactions in his description. However, he never could include interactions to his theory and that was reflected in the fact that his predictions significantly differed from the experiments, thus, it opened the path for alternative formulations of statistical mechanics.

1.1.2. Gibbs entropy

For the sake of providing an alternative foundation of statistical mechanics, Gibbs proposed in his work [25] a form of study statistical mechanics, where probability densities and *ensembles* play the main role. In his development, ensembles were infinite sets of macroscopically identical systems, each represented by a correspondent microstate, being compatible with a macrostate, and the probabilities appeared when he decided to explain the state of the system through a probability density ρ on its phase space.

$$S_{\mathcal{G}}(\rho) = -k \int \rho(X) \log \rho(X). \tag{1.3}$$

This particular identification of the state of the system and the success of this theory, lead to link the *Gibbs entropy* with the thermodynamic entropy. However, this particular identification, as pointed by Goldstein et. al. [33, 34], is not correct. To understand the point of Goldstein et. al., note that first, the Gibbs entropy is computed by a function of N particles instead of 1 particle as in Boltzmann's idea, and second, Gibbs entropy is a constant of motion. That is, if we write ρ_t for the evolution on densities induced by the motion on phase space, we have that $S_{\mathcal{G}}(\rho_t)$ is independent of t,

$$\frac{\mathrm{d}S_{\mathcal{G}}}{\mathrm{d}t} = 0. \tag{1.4}$$

It is frequently asked how this can be compatible with the second law if the entropy does not change. Ideas to answer these kind of questions have been proposed by Jaynes [35, 36], in which he argues that even though Gibbs entropy does not change over time, the distributions ρ_t for posterior measurements will lead to an entropy $S'_{\mathcal{G}}$ such that $S_{\mathcal{G}} \leq S'_{\mathcal{G}}$ and thus, the second law will be recovered. This position has been strongly criticised by S. Goldstein et. al. [33, 34], where they stated that the real thermodynamic entropy is the one provided by Boltzmann and that the Gibbs entropy is not even an entity of the right sort to describe what should be understood as the thermodynamic entropy, because in Gibbs entropy, ρ corresponds to a probability distribution, that is, a probability of an ensemble of systems, and it is not a function on phase space, a function of the actual state X of an individual system.

One could wonder what is so attractive about the Gibbs entropy since it does not provide a correct interpretation of thermodynamics. The answer to this is simple: Gibbs' approach is simple, elegant and produce the correct answers when predicting thermodynamic quantities. In Gibbs' perspective, the idea of assigning a probability density function on the phase space $\Omega(\Lambda)$, not only allows us to compute expected values of observables, but it also let us understand the state of a subsystem by simply computing the corresponding marginal probability from the probability density describing the whole system of N particles. Moreover, Gibbs' achievement [25] was to realize that there is a canonical measure over the phase space that allows us to conveniently define our probabilities. The measure that allows us to do that is given by the Darboux theorem [37, 38], and is known as the measure of the *symplectic form*. Specifically, when Gibbs looked at the problem of the gas with constant energy and fixed volume, he assigned equal probabilities to all possible systems that look macroscopically like X. In his own words [25]:

> *All microstates accessible to an isolated system are equally probable, because there is no evidence that certain microstate should be more probable than others.*

Namely, Gibbs considered that whenever a macroscopic system is at equilibrium, every state compatible with the constraints of the system has to be equally available (likely) compared to the others. Mathematically this translates into the choice of a constant density function, called the *micro-canonical* ensemble.

The two notions of Gibbs and Boltzmann entropy are parallel to two notions of thermal equilibrium, notions that are described by Goldstein et. al. [34] as the *ensemblist* and the *individualist* point of view. In the view of the ensemblist, a system is in thermal equilibrium if and only if its phase point X is random with the appropiate distribution, such as the micro-canonical distribution. In contrast, in the individualist view, a system is in thermal equilibrium if and only if its phase point X lies in a certain subset Γ_{Eq} of phase space. These two positions allow us to compare in a clearer way the main differences between Boltzmann and Gibbs entropy for classical systems. However, our world is quantum and the arguments we have discussed so far are classic. In the next section we are going to discuss how Gibbs can be implemented in the quantum case.

1.2. The quantum case

In quantum mechanics, a system is described by a vector in Hilbert space $\mathcal{H}$ and its evolution is generated by a Hamiltonian operator $\hat{H}$. Hence, when we consider a macroscopic quantum system with Hilbert space $\mathcal{H}$, we can think on translate Gibbs' ideas in a quantum mechanical context. First, any description of state of knowledge in a quantum mechanical systems has to be in terms of the maximum available information [39]. That is, the quantum systems has to be in a quantum pure state $|\psi_1\rangle$ with probability p_1 or it may be in the state $|\psi_2\rangle$ with probability p_2, etc. All the alternatives $|\psi_i\rangle$ are not necessarily mutually orthogonal, but each may be expanded in terms of a complete orthonormal set of functions $|\phi_k\rangle$

$$|\psi_i\rangle = \sum_k c_{ki} |\phi_k\rangle . \tag{1.5}$$

This state of knowledge is interpreted by a point P_i with coordinates c_{ki}. At each point P_i, we place a weight p_i, such that at the end we have a collection of weights p_i assigned to a state. Since each of the possible wave functions is normalised to unity,

$$\langle\psi_i, \psi_i\rangle = \int |\psi_i|^2 d\tau = 1, \tag{1.6}$$

we have that

$$\sum_k |c_{ik}|^2 = 1, \tag{1.7}$$

and all points P_i lie over the unit hypersphere. If each of the possible states $|\psi_i\rangle$ satisfies the same Schrödinger equation, then as time goes on, the dynamics of the points P_i can be seen as a rigid rotation hypersphere. Hence, the measure over which we assign the probabilities have to be an invariant measure over unitary transformations and a uniform measure over the sphere. This measure is known as the *Haar measure*. Moreover, p_i are not in general the probabilities of mutually exclusive events. In quantum mechanics, if a state is known to be in the state $|\psi_i\rangle$, then the probability of finding it upon measurement $|\psi_j\rangle$ is given by $| \langle\psi_i, \psi_i\rangle |^2$. Therefore, the probabilities p_i refer to independent mutually exclusive events only when the states $|\psi_i\rangle$ are mutually orthogonal states. Since nothing assures that we only work with orthogonal states, it is convenient to define an entropy that takes this into account

$$S_{vN}(\hat{\rho}) = -k\,\mathrm{Tr}(\hat{\rho}\log\hat{\rho}), \tag{1.8}$$

where $\hat{\rho}$ corresponds to the density matrix. This expression is known as the *quantum Gibbs entropy* or simply as the *von Neumann entropy* [40]. Note that this definition of entropy assigns zero entropy to any quantum pure state, and that similarly as in the classical case, this entropy is a constant in time, meaning that it can not account for the second law of thermodynamics.

This ensemblist viewpoint has been criticised [33, 34, 41, 42] to rely on subjective randomness and ensemble averages that in certain occasions do not have a clear physical meaning. However, in the last decades, there has been a wave of works dedicated to quantum thermalisation [1–7], which often are connected with the key words *eigenstate thermalization hypothesis* (ETH), and *canonical typicality*, in which an individualist viewpoint is addressed. A common factor in all these works is that an individual, closed, macroscopic quantum system in a pure state $|\psi(t)\rangle$, that evolves unitarily, under conditions usually satisfied, will behave very much as one would expect from a system in thermal equilibrium to behave [34]. In the next section, we will introduce the ideas behind typicality and we will dive in the details of it.

1.3. Canonical typicality

The purpose of this section is to explain why in quantum systems, the ensemblist ideas are not necessary to explain thermalisation, and an individualist view point can be used instead.

Consider a large quantum mechanical system, we will call it *the universe*, that can be decomposed in two parts, the system S and its environment E, where the dimension of the environment $d_{\mathcal{E}}$ is considered to be much larger than the dimension of the system $d_{\mathcal{S}}$. Now, suppose the universe has to obey some global constraint R, which translates into the choice of a subspace of the total Hilbert space, say

$$\mathcal{H}_R \subset \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{E}}, \tag{1.9}$$

where the dimension of $\mathcal{H}_R$ is denoted by d_R. When we deal with the standard approach in statistical mechanics, the restriction is imposed over the total energy. However, as Popescu et. al. emphasise in [1, 4], this restriction can be completely arbitrary and not necessarily referring to the energy.

Let $\mathcal{E}_R$ be the equiprobable state in $\mathcal{H}_R$,

$$\mathcal{E}_R = \frac{\mathbb{I}_R}{d_R}, \tag{1.10}$$

where $\mathbb{I}_R$ corresponds to the projector operator on $\mathcal{H}_R$, and $\mathcal{E}_R$ correspond with the maximally mixed state in $\mathcal{H}_R$, because it corresponds to the state of maximum ignorance in $\mathcal{H}_R$.

We define Ω_S, the canonical state of the system corresponding to the restriction R, as the quantum state of the system when the universe is in the equiprobable state $\mathcal{E}_R$. The canonical state of the system Ω_S is therefore obtained by tracing out the environment in the equiprobable state of the universe:

$$\Omega_S = \mathrm{Tr}_E\, \mathcal{E}_R. \tag{1.11}$$

Now, instead of considering the universe in the equiprobable state $\mathcal{E}_R$, we consider the universe to be in a random pure state $|\phi\rangle \in \mathcal{H}_R$. In that case, the system will be described by its reduced density matrix

$$\rho_S = \mathrm{tr}_B(|\phi\rangle\langle\phi|). \tag{1.12}$$

In the spirit of Boltzmann's ideas one could think that at an individual level, thermalisation may appear as consequence that the vast majority of states have the property that they equilibrate. As we will see, here we address similar ideas from a quantum point of view. Thus, we ask ourselves, "how different is ρ_S form the canonical state Ω_S?". The answer to this question is addressed by Popescu et. al. in [1,4], which states that ρ_S is very close to Ω_S for almost every pure state compatible with the constraint R. That is, for almost every pure state of the universe, the system behaves as if the universe were actually in the equiprobable mixed state $\mathcal{E}_R$.

To formally express canonical typicality, it is necessary to first define a notion of distance between the states ρ_S and Ω_S, as well as a measure over which pure states $|\phi\rangle$ are defined.

We define the trace distance between ρ_S and the canonical state Ω_S, by $||\rho_S - \Omega||_1$, this distance is explicitly calculated by

$$||\rho||_1 = \mathrm{Tr}\,|\rho| = \mathrm{Tr}\left(\sqrt{\rho^\dagger \rho}\right). \tag{1.13}$$

Consider $|\phi\rangle$ to be a pure state in $\mathcal{H}_R$, with respective dimension d_R. As the state is normalized ($\langle\phi|\phi\rangle = 1$) we know that the pure state $|\phi\rangle$ lives in a $(2d_R - 1)$-dimensional real sphere. Thus, the states the state $|\phi\rangle$ lives over the sphere surface of d_R dimensions. Hence, if we randomly sample pure states, we will have to sample them with the previously discussed Haar measure, which is the measure that is invariant under unitary transformations.

Once defined the notion of distance as well as the measure over which the states are sampled, we are ready to announce the theorem of canonical typicality.

Theorem 1.3.1 (Theorem of Canonical Typicality [1, 4].). *For a random chosen state, sampled with the Haar measure, $|\phi\rangle \in \mathcal{H}_R \subset \mathcal{H}_S \otimes \mathcal{H}_B$ and arbitrary $\varepsilon > 0$ the distance between the reduced density matrix $\rho_S = \mathrm{Tr}_E(|\phi\rangle\langle\phi|)$ and the canonical state $\Omega_S = \mathrm{Tr}_E\,\mathcal{E}_R$ is given probabilistically by:*

$$\mathrm{Prob}\left(\|\rho_S - \Omega_S\|_1 \geq \eta\right) \leq \eta', \tag{1.14}$$

where

$$\eta = \varepsilon + \sqrt{\frac{d_S}{d_E^{\mathrm{eff}}}}, \quad \eta' = 2\exp\left(-C d_R \varepsilon^2\right), \tag{1.15}$$

with

$$C = \frac{1}{18\pi^3}, \quad d_E^{\mathrm{eff}} = \frac{1}{\mathrm{Tr}\,\Omega_E^2} \geq \frac{d_R}{d_S}, \quad \Omega_E = \mathrm{Tr}_S\,\mathcal{E}_R. \tag{1.16}$$

Note that η and η' are small quantities, thus, we can assert that whenever $d_E^{\mathrm{eff}} \ggg d_S$ and $d_R\varepsilon^2 \ggg 1$, every reduced state will be close to its correspondent canonical state. What this results tells us is that probabilistically speaking, if the dimension of the accessible space (d_R) is large enough, we will have that for the overwhelming majority of choices of random pure states, will have almost certainly that every subsystem, with small enough dimension, will be indistinguishable from the canonical state. Moreover, Popescu et. al [4] found a bound to the average differences between the state of the system (the reduced state) and the canonical state. Explicitly, using the levy lemma, they are able to show

$$\langle \|\rho_S - \Omega_S\|_1 \rangle \leq \sqrt{\frac{d_S}{d_E^{\mathrm{eff}}}} \leq \sqrt{\frac{d_S^2}{d_R}}. \tag{1.17}$$

This bound tells us that most of pure states constrained to a global restriction have the property that when we look at the local state of the system it seem to behave as the thermal state.

Despite this result explains very well the reason why by randomly choosing a state $|\phi\rangle$ over the Haar measure, it coincides with the canonical state in almost all cases, it does not explain the way a state out of equilibrium (an atypical state), reaches equilibrium. The reason for that is that no particular evolution was considered here and only probabilistic arguments were used. This means that typicality is a kinematic description of thermalisation. However, because almost all states of the universe have the property that the local states are approximately the canonical state, we anticipate that most evolutions will quickly carry a state in which the system is not thermalised to one in which it is. Namely, in the next section we will see how from a typicality viewpoint, Linden et. al. [18] were able to show that thermalisation can occur in a system reliant on a unitary dynamic.

1.4. Evolution towards equilibrium.

Typicality is kinematic result, meaning that is only valid for a given time and a given state, and were unitary evolution does not play a role. Specifically, we are interested in states that are atypical, in the sense that are states that locally differ from the canonical state. As pointed out by Linden et. al. [18], to prove thermalisation is a much complicated problem since a closer look of the elements present in thermalisation, shows that equilibration, environment state independence, system state independence and the Boltzmannian form of the state of equilibrium are needed to assure that thermalisation has taken place. Specifically, the result obtained by Linden et. al. in [18] addressed only the first two elements, showing that, equilibrium can be understood as a local universal property of quantum systems. It is important to stress that when we refer to equilibrium we do not necessarily refer to thermal equilibrium; indeed, the equilibrium state can be an arbitrary state with only the property that it does not change over time.

To understand how is possible to prove that equilibrium appear as a “typical behaviour” in quantum systems. It is important to first define some concepts that we have used before but that we consider are important to make explicit.

Universe, system and environment: For us the universe will always refer to a large quantum system living in a Hilbert space $\mathcal{H}$. As before, our universe is always decomposed in two, and in this decomposition we refer to the system S as a small part of the total Hilbert space. The remaining, we will call it the environment. Explicitly, we will always decompose the Hilbert space of the universe as a tensor product of the Hilbert space of the system and the environment, $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$, where d_S and d_E are the respective dimension of the system and the environment.

Note that neither the environment nor the system have been provided with any special property, meaning that for this formulation, the system could be a single particle or even a section of a lattice.

For the sake of proving the result in [18], we also define the Hamiltonian of the universe as:

Definition 1.4.1 (Hamiltonian)**.** *The evolution of the universe will be governed by a Hamiltonian given by*

$$\hat{H} = \sum_k E_k \left|E_k\right\rangle \left\langle E_k\right|, \tag{1.18}$$

with $|E_k\rangle$ *the eigenstate in the energy basis with energy* E_k*. Where the main required assumption is that the Hamiltonian has non-degenerate energy gaps.*

Expressing this condition in a more explicit way, it is said that a Hamiltonian has no-degenerate energy gaps if any non-zero difference of eigenvalues of energy determine the two energy values involved. That is, for any four eigenstates with energy E_k, E_ℓ, E_m, E_n, satisfy that if $E_k - E_\ell = E_m - E_n$, then $m = n$ and $k = \ell$, or $k = m$ and $\ell = n$.

Notice that the restriction imposed to the Hamiltonian is an extremely natural constraint, because all Hamiltonians that lack of symmetries have non-degenerate energies, so we talk about a set of Hamiltonians with measure 1 that fulfils this condition.

Notation: We will work here with pure time dependent states of the universe, states that will be represented by $|\Psi(t)\rangle$ with a time dependent density matrix given by $\rho(t) = |\Psi(t)\rangle \langle\Psi(t)|$.

As the state of the system at a time t can be found by tracing out the environment, that is, $\rho_S(t) = \mathrm{Tr}_E\, \rho(t)$,identically, we define the state of the environment as $\rho_E(t) = \mathrm{Tr}_S\, \rho(t)$.

It is convenient to define the transient states of the universe, or the time averaged state ω as

$$\omega = \langle\rho(t)\rangle_t = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau \rho(t) dt. \tag{1.19}$$

This definition allows us to also define ω_s and ω_E as the time averaged state of the system and the environment respectively. Finally, we introduce the concept of the effective dimension of a mixed state ρ:

$$d^{\mathrm{eff}}(\rho) = \frac{1}{\mathrm{Tr}\left(\rho^2\right)}. \tag{1.20}$$

The meaning of this effective dimension is how many states contribute to the mixture, carrying the probabilistic weight of different states, and different than the support of an operator in the Hilbert space, it is a continuous measure.

With the concepts aforementioned, Linden et. al. [18] are able to show that every pure state of a quantum universe, composed by a large number of eigenstates of energy, and that evolves under an arbitrary Hamiltonian, is such that every small system will equilibrate. The reason of consider the global state to have many eigenstates of energy is because if there are many eigenstates, we can assure that there will be a large quantity of changes throughout the evolution of the system. The notion of evolving through many states can be mathematically encapsulated via the effective dimension of the time average state $\omega = \langle \rho(t) \rangle_t$, and the connection between this and the number of eigenstates is with ease seen by expanding a time dependent state $|\Psi(t)\rangle$ as

$$|\Psi(t)\rangle = \sum_k c_k e^{-iE_k t} |E_k\rangle \tag{1.21}$$

where $\sum_k |c_k|^2 = 1$ and hence

$$\rho(t) = \sum_{k,l} c_k c_l^* e^{-i(E_k - E_l)t} |E_k\rangle \langle E_l| , \tag{1.22}$$

that can be expanded and written as

$$\begin{aligned}\rho(t) &= \underbrace{\sum_n \|c_n\|^2 |E_n\rangle \langle E_n|}_{\omega} + \underbrace{\sum_{m \neq n} c_n c_m^* |E_n\rangle \langle E_m| e^{-it(E_n - E_m)}}_{\lambda(t)} \\ &= \omega + \lambda(t).\end{aligned} \tag{1.23}$$

For the case of non-degeneracy of the energy levels, we have that the cross-terms vanish, that is

$$\omega = \langle \rho(t) \rangle_t = \sum_k |c_k|^2 |E_k\rangle \langle E_k| , \tag{1.24}$$

which leads us to

$$d^{\text{eff}}(\omega) = \frac{1}{\text{Tr}\,(\omega^2)} = \frac{1}{\sum_k |c_k|^4}. \tag{1.25}$$

In the same way as in typicality, we are going to ask ourselves about the distance between $\rho_S(t)$ and $\omega_S = \langle \rho_S(t) \rangle_t$. To do this, we first compute the difference between $\rho_S(t)$ and ω_S in terms of the energy eigenstates as

$$\rho_S(t) - \omega_S = \sum_{m \neq n} c_m c_n^* e^{-i(E_m - E_n)t}\, \text{Tr}_E\, |E_m\rangle \langle E_n| . \tag{1.26}$$

Since in general we know that $\rho_S(t)$ fluctuates around the state ω_S, it is evident that the distance between them will change over time. Thus, we will be interested in the time average of the trace distance $\langle ||\rho_S(t) - \omega_S||_1 \rangle_t$. The value this average takes will tell us about where the system is spending most of its time. In other words, $\langle ||\rho_S(t) - \omega_S||_1\rangle_t$ will be small when the system equilibrates to ω_S. To be able to prove what is announced as the *Theorem 1* in [18] it is useful to relate the trace distance to the square of the Hilbert-Schmidt distance using a standard bound provided in [43]

$$||\rho_1 - \rho_2||_1 = \frac{1}{2}\,\mathrm{Tr}_S\sqrt{(\rho_1-\rho_2)^2} \leq \frac{1}{2}\sqrt{d_S\,\mathrm{Tr}_S\,(\rho_1-\rho_2)^2}. \tag{1.27}$$

Which combined with the concavity of the square-root function, yields:

$$\langle ||\rho_S(t) - \omega_S||_1\rangle_t \leq \sqrt{d_S\left\langle \mathrm{Tr}_S\left[\rho_S(t)-\omega_S\right]^2\right\rangle_t}, \tag{1.28}$$

that provides us the bound we need to proof the theorem. Now using (1.26) we write:

$$\left\langle \mathrm{Tr}_{\mathcal{S}}\left[\rho_{\mathcal{S}}(t)-\omega_{\mathcal{S}}\right]^2\right\rangle_t = \sum_{m\neq n}\sum_{k\neq l}\mathcal{T}_{klmn}\,\mathrm{Tr}_{\mathcal{S}}\left(\mathrm{Tr}_E\left|E_k\right\rangle\left\langle E_l\left|\mathrm{Tr}_E\right|E_m\right\rangle\left\langle E_n\right|\right), \tag{1.29}$$

where $\mathcal{T}_{klmn} = c_k c_l^* c_m c_n^* e^{-i(E_k-E_l+E_m-E_n)t}$. We compute the time average taking into account that the Hamiltonian has non-degenerate energy gaps, thus, we find that

$$\begin{aligned}
\left\langle \mathrm{Tr}_S\left[\rho_S(t)-\omega_S\right]^2\right\rangle_t &= \sum_{k\neq l}|c_k|^2|c_l|^2\,\mathrm{Tr}_S\left(\mathrm{Tr}_E\left|E_k\right\rangle\left\langle E_l\left|\mathrm{Tr}_E\right|E_l\right\rangle\left\langle E_k\right|\right)\\
&= \sum_{k\neq l}|c_k|^2|c_l|^2\sum_{ss'bb'}\langle sb|E_k\rangle\langle E_l|s'b\rangle\langle s'b'|E_l\rangle\langle E_k|sb'\rangle\\
&= \sum_{k\neq l}|c_k|^2|c_l|^2\sum_{ss'bb'}\langle sb|E_k\rangle\langle E_k|sb'\rangle\langle s'b'|E_l\rangle\langle E_l|s'b\rangle\\
&= \sum_{k\neq l}|c_k|^2|c_l|^2\,\mathrm{Tr}_E\left(\mathrm{Tr}_S\left|E_k\right\rangle\left\langle E_k\left|\mathrm{Tr}_S\right|E_l\right\rangle\left\langle E_l\right|\right)\\
&= \sum_{k\neq l}\mathrm{Tr}_E\left[\mathrm{Tr}_S\left(|c_k|^2\left|E_k\right\rangle\left\langle E_k\right|\right)\mathrm{Tr}_S\left(|c_l|^2\left|E_l\right\rangle\left\langle E_l\right|\right)\right]\\
&= \mathrm{Tr}_E\,\omega_E^2 - \sum_k|c_k|^4\,\mathrm{Tr}_S\left[\left(\mathrm{Tr}_E\left|E_k\right\rangle\left\langle E_k\right|\right)^2\right]\\
&\leq \mathrm{Tr}_E\,\omega_E^2,
\end{aligned} \tag{1.30}$$

where $\omega_E = \mathrm{Tr}_S\,\omega$. To obtain a further bound, we invoke weak sub-additivity of the Reényi entropy [44]

$$\mathrm{Tr}\left(\omega^2\right) \geq \frac{\mathrm{Tr}_E\left(\omega_E^2\right)}{\mathrm{rank}\left(\rho_S\right)} \geq \frac{\mathrm{Tr}_E\left(\omega_E^2\right)}{d_S}. \tag{1.31}$$

Hence, combining (1.28), (1.30) and (1.31) we get

$$\langle ||\rho_S(t) - \omega_S||_1\rangle_t \leq \frac{1}{2}\sqrt{d_S\,\mathrm{Tr}_E\left(\omega_E^2\right)} \leq \frac{1}{2}\sqrt{d_S^2\,\mathrm{Tr}\left(\omega^2\right)}. \tag{1.32}$$

By taking the definition of effective dimension, we get the main result shown in [18]

$$\langle ||\rho_S(t) - \omega_S||_1 \rangle_t \leq \frac{1}{2}\sqrt{\frac{d_S}{d^{\text{eff}}\left(\omega_E\right)}} \leq \frac{1}{2}\sqrt{\frac{d_S^2}{d^{\text{eff}}(\omega)}}. \tag{1.33}$$

As we can see, the result obtained by Linden et. al. tells us that the vast majority of quantum systems, in which the dynamics of the universe is governed by a Hamiltonian with no gaps, will spend most of its time close to its equilibrium state independently of its initial state. Note that this result is not necessarily considering that the state of equilibration will coincide with the canonical state, but since we expect to have thermal typicality in the system, that state of equilibrium will coincide with the thermal equilibrium.

1.5. Consequences of typicality

We presented typicality as an alternative mechanism to understand thermalisation in large quantum systems, and we saw how these ideas let us explain equilibration in large quantum systems. The purpose of this section will be to point some of the consequences within typicality.

Consider two different orthogonal pure states living in the same Hilbert subspace ($|E_n\rangle, |E_m\rangle \in \mathcal{H}_R$), the Hilbert subspace associated with the global restriction R. From typicality, we know that the each of the reduced states of $|E_n\rangle, |E_m\rangle$ approximately leads to the canonical state, that is,

$$\text{Tr}_{\mathcal{E}} |E_n\rangle \langle E_n| \approx \text{Tr}_{\mathcal{E}} |E_m\rangle \langle E_m| \approx \Omega_{\mathcal{S}}. \tag{1.34}$$

Consider a third state $|\Psi\rangle$ to be a generic linear combination of $|E_n\rangle$ and $|E_m\rangle$. In this case we will have that the density matrix associated with the third state is:

$$\begin{aligned} \rho &= |\Psi\rangle \langle\Psi| \\ &= \|c_n\|^2 |E_n\rangle \langle E_n| + \|c_m\|^2 |E_m\rangle \langle E_m| \\ &+ c_n c_m^* |E_n\rangle \langle E_m| + c_m c_n^* |E_m\rangle \langle E_n| . \end{aligned} \tag{1.35}$$

If we take the partial trace of the equation (1.35), we will end up with the state of the system,

$$\begin{aligned} \Omega(E) \approx \text{Tr}_E\, \rho &= \text{Tr}_E |\Psi\rangle \langle\Psi| \\ &= \|c_n\|^2 \text{Tr}_E |E_n\rangle \langle E_n| + \|c_m\|^2 \text{Tr}_E |E_m\rangle \langle E_m| \\ &+ c_n c_m^* \text{Tr}_E |E_n\rangle \langle E_m| + c_m c_n^* \text{Tr}_E |E_m\rangle \langle E_n| . \end{aligned} \tag{1.36}$$

Notice that the way we constructed the state $|\Psi\rangle$, makes the non-crossed terms in (1.36) to approximately coincide with $\mathrm{Tr}_E |E_n\rangle \langle E_n| = \mathrm{Tr}_E |E_m\rangle \langle E_m| \approx \Omega(E)$, with $\Omega(E)$ the canonical state. Thus, the equation (1.36) can be written as:

$$\Omega(E) \approx \rho_S = \Omega(E) + c_n c_m^* \, \mathrm{Tr}_E |E_n\rangle \langle E_m| + c_m c_n^* \, \mathrm{Tr}_E |E_m\rangle \langle E_n| \,. \tag{1.37}$$

In order to get the equality, the cross terms in (1.37) have to approximately vanish. This property of vanishing partial traces of exterior products of states is what we name *ultra-orthogonality*. Explicitly, that is,

$$\mathrm{Tr}_E |E_n\rangle \langle E_m| = \mathrm{Tr}_E |E_m\rangle \langle E_n| \approx 0. \tag{1.38}$$

Notice that this property appears naturally by just using the results of typicality. Also, the property of getting vanishing partial traces over the crossed terms could explain an alternative path to equilibrium as an instant phenomenon. To see this, we can replace the condition (1.38) in equation (1.26) and see that:

$$\mathrm{Tr}_{\mathcal{E}} \, \rho(t) \equiv \rho_{\mathcal{S}} \approx \omega_{\mathcal{S}}. \tag{1.39}$$

Notice that whenever ultra-orthogonality holds, temporal averages are not needed and equilibration will appear as an instantaneous phenomenon. Inspired on this interesting phenomena, we decided to explore ultra-orthogonality for the special case in which the terms in (1.38) of the left hand side are exactly equal to zero, and thus implying the equality in equation (1.39). This means that the correspondent reduced state of a fully interacting universe will be immediately constant. The idea of this work is to explore if it is possible to find Hilbert subspaces in which ultra-orthogonality holds, and more importantly, to find out how large these Hilbert subspaces can be. Specifically we will be interested in the question "what is the largest Hilbert subspace in which ultra-orthogonality will hold?" In the next chapters we will tackle this question for the case of systems that their Hamiltonians can be mapped, under appropriate approximations (or transformations), to Hamiltonians which are quadratic in fermionic operators.

Chapter 2

Theoretical background

At the end of the last chapter we announced that the problem we want to study in this document is that of ultra-orthogonality. Particularly, in the previous chapter, we showed how ultra-orthogonality is connected to the problem of equilibration as an instantaneous phenomenon. Since this problem is quite general, we focus our research to study the particular case in which ultra-orthogonality holds exactly for a particular kind of fermionic systems. For that reason, the purpose of this chapter is to provide the necessary theoretical background to study ultra-orthogonality for the aforementioned case.

In this chapter we introduce the concepts of quasi-free fermionic models on a lattice and present Majorana fermions to define the fermionic covariance matrix. We show how this formalism can be used to develop some standard calculations on the diagonalisation of the Hamiltonian of the one-dimensional XY model, and we will discuss how is possible to treat excited states as well as local reduced states in this model. Since part of this work is dedicated to show an explicit connection between a special case of fermionic systems and code theory, the second part of this chapter will be devoted to introduce some concepts of code theory that will allows us to make the explicit connection between code theory and fermionic system in the third chapter.

2.1. Fermionic quadratic Hamiltonian

In many areas of physics one has to deal with solving quantum many body problems. This is often a computationally difficult if not an impossible task. However, the cases that can be analytically solved are well-known, and have been a subject of study [45–57].

It has been found that a wide class of complicated Hamiltonians with many-body interactions can often be mapped onto Hamiltonians that are quadratic in annihilation and creation operators and have the generic form [51]

$$\hat{H} = \sum_{ij} C_{ij}\hat{a}_i^\dagger \hat{a}_j + \sum_{ij}\left(A_{ij}\hat{a}_i^\dagger \hat{a}_j^\dagger + \text{h.c.}\right), \tag{2.1}$$

where i, j run from 1 to N, the number of modes in the system and $\hat{a}_i$, $\hat{a}_i^\dagger$ are fermionic annihilation and creation operators that satisfy the canonical anti-commutation relations (CAR) [58]

$$\{\hat{a}_k, \hat{a}_l\} = \left\{\hat{a}_k^\dagger, \hat{a}_l^\dagger\right\} = 0, \quad \left\{\hat{a}_k, \hat{a}_l^\dagger\right\} = \delta_{kl}. \tag{2.2}$$

A convenience when working with these kind of Hamiltonians is that they can be diagonalised via a so-called Bogoliubov - Valantin transformation [59], also known as a canonical transformation, which maps fermionic creation and annihilation operators to creation and annihilation operators of non-interacting quasi-particles [59, 60]. Explicitly, the transformation looks like

$$\begin{aligned} \hat{a}_i &\mapsto \alpha\hat{q}_i + \kappa_i\hat{q}_i^\dagger, \\ \hat{a}_i^\dagger &\mapsto \bar{\alpha}_i\hat{q}_i^\dagger + \bar{\kappa}_i\hat{q}_i. \end{aligned} \tag{2.3}$$

where α_i, κ_i are complex numbers such that preserves the canonical anti-commutation relations given by (2.2) for $\hat{q}$, $\hat{q}^\dagger$. This relation can also be expresses as a condition over γ_i, κ_i,

$$\gamma_i^2 + \kappa_i^2 = 1, \tag{2.4}$$

and

$$\{\hat{q}_k, \hat{q}_l\} = \left\{\hat{q}_k^\dagger, \hat{q}_l^\dagger\right\} = 0, \quad \left\{\hat{q}_k, \hat{q}_l^\dagger\right\} = \delta_{kl}. \tag{2.5}$$

The Bogoliubov-Valantin is relevant in many physics models because this transformation diagonalise many Hamiltonians; some examples of this are the Hubbard model, the BCS theory of superconductivity in the mean field or Hartree-Fock approximation, and certain solvable spin-chain models (After a Jordan-Wigner transformation) [55–58].

Hamiltonians with the generic form of (2.1) have the interesting property that not only the ground state but every eigenstate representing a certain number of excitations of quasi-particles, described by $\hat{a}$ and $\hat{a}^\dagger$, belong to the so-called class of fermionic Gaussian states, which is an interesting property, since it allows us to characterise them in terms of second order correlations, and the reason is because all the higher moments factorize as stated in Wick's theorem [50, 61]. An equivalent but convenient characterization of second order correlations are defined in terms of Majorana fermions as we will see bellow .

2.2. Majorana Fermions

Majorana fermions are represented in terms of $2N$ hermitian operators

$$\hat{\gamma}_j = \hat{a}_j^\dagger + \hat{a}_{j+N}, \quad \hat{\gamma}_{j+N} = (-i)\left(\hat{a}_j^\dagger - \hat{a}_j\right), \tag{2.6}$$

where these operators are analogous to coordinate and momentum operators for bosonic modes, and for each fermion labelled by j of the original system we define two operators above. The canonical Fermi-Dirac commutation relation takes the form

$$\{\hat{\gamma}_k, \hat{\gamma}_l\} = 2\delta_{kl}. \tag{2.7}$$

The algebra generated by the operators $\{\hat{\gamma}_i\}$ is known as the Clifford algebra and is denoted by $\mathcal{C}_{2N}$[1]. When we change from the Fermionic operators $a^T := (\hat{a}_,\hat{a}_2, \ldots, \hat{a}_N, \hat{a}_1^\dagger, \hat{a}_2^\dagger, \ldots, \hat{a}_N^\dagger)$ to Majorana operators $\gamma^T := (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_N, \hat{\gamma}_{N+1}, \ldots, \hat{\gamma}_{2N})$, it is convenient to define the Fermionic covariance matrix which will fully characterise Gaussians states.

2.3. Fermionic Covariance matrix

As we mentioned before, Gaussian states are completely characterised by its second moments [50, 61], that is, Gaussian states have a density matrix ρ [62],

$$\rho = \frac{1}{Z} \cdot \exp\left[-\frac{i}{4}\hat{\gamma}^T G \hat{\gamma}\right], \tag{2.8}$$

with $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_{2N})$, the vector of Majorana operators (2.6), Z a normalization constant and G real skew-symmetric $2N \times 2N$ matrix. Since G is a skew-symmetric matrix, it can always be brought to the block diagonal form

$$OGO^T = \begin{pmatrix} 0 & -\tilde{B} \\ \tilde{B} & 0 \end{pmatrix} \quad \text{with} \quad O \in \mathrm{SO(2N)}, \tag{2.9}$$

where $\tilde{B}$ is diagonal, with eigenvalues that we denote by $\tilde{\beta}_k$. The right hand side of (2.9) is known as the Williamson form of the skew-symmetric matrix G, and $\tilde{\beta}_k$ are the Williamson

[1]The orthogonal group in $2N$ dimensions $O(2N)$ preserves the Clifford algebra, hence, the canonical canonical Fermi-Dirac commutation relations of fermionic operators

eigenvalues of G [63].

It is convenient to characterise second order correlations in terms of the so-called *fermionic covariance matrix* (FMC), whose entries are

$$\Gamma_{kl} = \frac{i}{2}\,\mathrm{Tr}\left(\rho\left[\hat{\gamma}_k, \hat{\gamma}_l\right]\right), \tag{2.10}$$

where $[\hat{\gamma}_k, \hat{\gamma}_l] := \hat{\gamma}_k\hat{\gamma}_l - \hat{\gamma}_l\hat{\gamma}_k$. Thus, we can bring this anti-symmetric matrix to its block diagonal form, via a canonical transformation, as

$$\tilde{\Gamma} = O\Gamma O^T = \begin{pmatrix} 0 & -\operatorname{diag}(\lambda_i) \\ \operatorname{diag}(\lambda_i) & 0 \end{pmatrix}, \tag{2.11}$$

where $\lambda_k = \tanh(\tilde{\beta}_k/2)$, for $k = 1, 2\ldots, N$ [63], which determines the connection between the matrix G in (2.9) and the FMC Γ. The Williamson eigenvalues are $\lambda_k = n_k - 1/2$, with n_k the fermion occupation number of the normal mode labelled by k.

The equivalence between the special orthogonal group in $2N$ dimensions ($SO(2N)$) and the Fermionic Gaussian states, leads to an interesting property about states describing multi-particles excitations. If $|0\rangle$ is the ground state of some Hamiltonian, with annihilation operators $\hat{a}_i$ in a given quasi-particle basis, then $\hat{a}_i^\dagger|0\rangle = \hat{c}_{2i}|0\rangle$. Meaning that if any multi-particle state of this kind is obtained from the ground state $|0\rangle$ through some transformation, such that preserves the canonical anti-commutation relation, the state will remain Gaussian. In other words, Gaussian states are preserved under any unitary transformation that preserves anti-commutation relations.

The fact that all eigenstates of the Hamiltonian in (2.1) are Gaussian is an important property, because it means that excited states can also be treated with the Covariance matrix formalism, and since we will be interested in the case of the excited states, it will be a property that we will exploit.

As an example of the afore-mentioned concepts, we will consider the case of the one-dimensional XY model and we will see how is possible to characterise this system through the Fermionic covariance matrix.

2.4. XY model.

The XY Hamiltonian model is a set of N spin 1/2 particles located on the sites of d-dimensional lattice. For the purpose of this document, whenever we refer to the XY model, we will have in mind the $1D$ XY model.

A chain of N spins where each spin is able to interact with its nearest neighbours in the X and Y component as well as an external magnetic field, will be described by the Hamiltonian of the form

$$H_{XY} = -\frac{1}{2}\sum_{l=0}^{N-1}\left(\frac{1+\gamma}{2}\hat{\sigma}_l^x\hat{\sigma}_{l+1}^x + \frac{1-\gamma}{2}\hat{\sigma}_l^y\hat{\sigma}_{l+1}^y + \lambda\hat{\sigma}_l^z\right), \tag{2.12}$$

where γ is so-called the anisotropy parameter and represents the difference between the strength of the XX interaction and the YY interaction in the spin space, λ is the intensity of the external magnetic field and

$$\hat{\sigma}_l^i = \mathbb{I}\otimes\cdots\otimes\mathbb{I}\otimes\underbrace{\hat{\sigma}^i}_{\text{site } l}\otimes\mathbb{I}\otimes\cdots\otimes\mathbb{I}, \tag{2.13}$$

where $\hat{\sigma}^i$ are Pauli matrices for $i = x, y, z$.

2.4.1. The spectrum

In order to find the spectrum of the of the XY model, it is necessary to perform three different transformations. These results are very standard and we present them to make our discussion self-consistent.

2.4.2. Jordan-Wigner transformation

The Jordan-Wigner transformation is an important transformation used mainly in Fermionic systems [64]. The Jordan-Wigner transformation provides a bridge between spins and fermions through a non-local transformation that maps spin operators onto fermionic creation and annihilation operators. Consider the next non-local transformation

$$\hat{a}_l = \left(\prod_{m<l}\hat{\sigma}_m^z\right)\hat{\sigma}_l^-, \quad \hat{\sigma}_l^- = \frac{\hat{\sigma}_l^x - i\hat{\sigma}_l^y}{2}, \tag{2.14}$$

where $\hat{a}_l$ represent spinless fermionic operators, and its canonical anticommutation relation (CAR) is given by [47]

$$\left\{\hat{a}_i^\dagger, \hat{a}_j^\dagger\right\} = \{\hat{a}_i, \hat{a}_j\} = 0, \quad \left\{\hat{a}_i^\dagger, \hat{a}_j\right\} = \delta_{i,j}. \tag{2.15}$$

Inverting the transformation we get

$$\begin{aligned}
\hat{\sigma}_l^z &= 1 - 2\hat{a}_l^\dagger\hat{a}_l,\\
\hat{\sigma}_l^x &= \left(\prod\nolimits_{m<l}\left(1-2\hat{a}_m^\dagger\hat{a}_m\right)\right)\left(\hat{a}_l^\dagger + \hat{a}_l\right),\\
\hat{\sigma}_l^y &= i\left(\prod\nolimits_{m<l}\left(1-2\hat{a}_m^\dagger\hat{a}_m\right)\right)\left(\hat{a}_l^\dagger - \hat{a}_l\right).
\end{aligned} \tag{2.16}$$

The terms of interaction in the Hamiltonian become

$$\begin{aligned}\hat{\sigma}^x_l\hat{\sigma}^x_{l+1} &= \left(\hat{a}^\dagger_l - \hat{a}_l\right)\left(\hat{a}^\dagger_{l+1} + \hat{a}_{l+1}\right),\\ \hat{\sigma}^y_l\hat{\sigma}^y_{l+1} &= -\left(\hat{a}^\dagger_l + \hat{a}_l\right)\left(\hat{a}^\dagger_{l+1} - \hat{a}_{l+1}\right),\end{aligned} \tag{2.17}$$

and the Hamiltonian of the XY model becomes,

$$H_{XY} = -\frac{1}{2}\sum_l \left[\left(\hat{a}^\dagger_{l+1}\hat{a}_l + \hat{a}^\dagger_l\hat{a}_{l+1}\right) + \gamma\left(\hat{a}^\dagger_l\hat{a}^\dagger_{l+1} - \hat{a}_l\hat{a}_{l+1}\right)\right] - \frac{\lambda}{2}\sum_l\left(1 - 2\hat{a}^\dagger_l\hat{a}_l\right), \tag{2.18}$$

with the boundary condition of $\hat{a}_N \equiv \hat{a}_1$, and where the term of $-\lambda N/2$ was ignored since it does not affect of the spectrum in the energy [47].

Note that we ended up with a Hamiltonian that only depends only on creation and annihilation operators and that has a similar shape of the Hamiltonian (2.1) presented at the beginning of the chapter.

2.4.3. Fourier transformation

If we consider periodic boundary conditions, that is, we identify the spin in site N with the spin in site 1, then, a Fourier transform can be applied to the operators $\hat{a}_l$ in the following way [47]

$$\hat{d}_k = \frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{a}_l e^{-i\phi_k l}, \quad \theta_k = \frac{2\pi}{N}k. \tag{2.19}$$

Note that the Fourier transformation is unitary, so the operators $\hat{d}_k$ are fermionic operators and will preserve the CAR.
In terms of $\hat{d}_k$ operators, the Hamiltonian (2.18) takes the form

$$H_{XY} = \sum_{k=-(N-1)/2}^{(N-1)/2}\left(-\lambda + \cos\phi_k\right)\hat{d}^\dagger_k\hat{d}_k + \frac{i\gamma}{2}\sum_{k=-(N-1)/2}^{(N-1)/2}\sin\phi_k\left(\hat{d}_k\hat{d}_{-k} + h.c\right), \tag{2.20}$$

where we have suppressed an additional term that is proportional to $1/N$ [56, 57], and the reason is because we are interested in the thermodynamic limit $N \to \infty$.

2.4.4. Bogoliubov -Valantin transformation

As mentioned in section 2.1, fermionic quadratic Hamiltonians can be easily diagonalised via a Bogoliubov-Valantin transformation over the operators $\hat{d}_k$

$$\tilde{d}_k = u_k\hat{d}^\dagger_k + iv_k\hat{d}_{-k}. \tag{2.21}$$

Since we want this transformation to preserve CAR, it is needed that $u_k^2 + v_k^2 = 1$, which implies that an appropriate parametrization will be $u_k = \cos(\psi_k/2)$ and $v_k = \sin(\psi_k/2)$, with

$$\cos\frac{\psi_k}{2} = \frac{-\lambda + \cos\phi_k}{\sqrt{(\lambda - \cos\phi_k)^2 + (\gamma\sin\phi_k)^2}}, \tag{2.22}$$

So finally our Hamiltonian will look as

$$H_{XY} = \sum_{-(N-1)/2}^{(N-1)/2} \tilde{\Lambda}(\theta_k)\tilde{d}_k^\dagger \tilde{d}_k, \tag{2.23}$$

with

$$\tilde{\Lambda}(\theta_k) := \sqrt{(\lambda - \cos\phi_k)^2 + (\gamma\sin\phi_k)^2}, \tag{2.24}$$

where the latter expression allow us to identify the critical regions of the model.

2.4.5. Fermionic covariance matrix for the XY model

Since we devote our work to study ultra-orthogonality, we have to be able to study properties of reduced density matrices of eigenstates of Hamiltonians quadratic in fermionic operators, and to be able to do that, it is important to characterise the covariance matrix of the XY model. In order to do this, we need to express the Hamiltonian (2.12) in terms of Majorana fermions using an analogous Jordan-Wigner transformation to the one used to diagonalise the XY Hamiltonian but into $2N$ Majorana fermions,

$$\hat{\gamma}_l = \left(\prod_{m<l} \hat{\sigma}_m^z\right)\hat{\sigma}_l^x, \quad \hat{\gamma}_{l+N} = \left(\prod_{m<l} \hat{\sigma}_m^z\right)\hat{\sigma}_l^y, \tag{2.25}$$

where again $l = 1, 2 \dots N-1$.

Note that the three following products:

$$\hat{\gamma}_l\hat{\gamma}_{l+N} = \left(\prod_{m<l} \hat{\sigma}_m^z\right)\left(\prod_{m<l} \hat{\sigma}_m^z\right)\hat{\sigma}_l^x\hat{\sigma}_l^y = i\hat{\sigma}_l^z, \tag{2.26}$$

$$\hat{\gamma}_{l+N}\hat{\gamma}_{l+1} = \left(\prod_{m<l} \hat{\sigma}_m^z\right)\hat{\sigma}_l^y\left(\prod_{m<l+1} \hat{\sigma}_m^z\right)\hat{\sigma}_{l+1}^x = \hat{\sigma}_l^y\hat{\sigma}_l^z\hat{\sigma}_{l+1}^x = i\hat{\sigma}_l^x\hat{\sigma}_{l+1}^x, \tag{2.27}$$

and

$$\hat{\gamma}_l\hat{\gamma}_{l+N+1} = \left(\prod_{m<l} \hat{\sigma}_m^z\right)\hat{\sigma}_l^x\left(\prod_{m<l+1} \hat{\sigma}_m^z\right)\hat{\sigma}_{l+1}^y = \hat{\sigma}_l^x\hat{\sigma}_l^z\hat{\sigma}_{l+1}^y = -i\hat{\sigma}_l^y\hat{\sigma}_{l+1}^y. \tag{2.28}$$

Coincide, up to constant factors, with the three terms in (2.12); then we can write the XY Hamiltonian as [54]

$$H_{XY} = \frac{i}{4} \sum_{\alpha,\beta=0}^{2N} \Omega_{\alpha\beta} \left[\hat{\gamma}_\alpha, \hat{\gamma}_\beta\right], \tag{2.29}$$

where Ω is the antisymmetric matrix of the form

$$\Omega = \left[\begin{array}{c|c} 0 & \tilde{\Omega} \\ \hline -\tilde{\Omega}^T & 0 \end{array}\right], \tag{2.30}$$

with

$$\tilde{\Omega} = \begin{pmatrix} \lambda & \frac{1-\gamma}{2} & 0 & 0 & \dots & 0 & \frac{1+\gamma}{2} \\ \frac{1+\gamma}{2} & \lambda & \frac{1-\gamma}{2} & 0 & \dots & 0 & 0 \\ 0 & \frac{1+\gamma}{2} & \lambda & \frac{1-\gamma}{2} & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \dots & \vdots & \vdots \\ \frac{1-\gamma}{2} & 0 & 0 & 0 & \dots & \frac{1+\gamma}{2} & \lambda. \end{pmatrix}. \tag{2.31}$$

Given that $\tilde{\Omega}$ is a circulant matrix, it can be diagonalised by means of a Fourier transformation. Therefore it can be written as

$$\tilde{\Omega}_{mn} = \frac{1}{N} \sum_{\theta_k \in (-\pi,\pi)} \omega\left(\theta_k\right) e^{\phi(\theta_k)} e^{i(m-n)\theta_k}. \tag{2.32}$$

where $\omega\left(\theta_k\right) = \omega\left(\theta_k\right)^* = \omega\left(-\theta_k\right), \phi\left(\theta_k\right) = -\phi\left(\theta_k\right)$ and are given by

$$\omega^2\left(\theta_k\right) := \left(\lambda - \cos\theta_k\right)^2 + \gamma^2 \sin^2\theta_k, \tag{2.33}$$

and

$$\phi\left(\theta_k\right) := \arctan\left(\frac{\lambda - \cos\theta_k}{-\gamma\sin\theta_k}\right). \tag{2.34}$$

The summation in (2.32) is understood over k with $-(N-1)/2 \leq k \leq (N-1)/2$, which is equivalent to $-\pi \leq \theta_k \leq \pi$. So defining the following functions

$$u_m^c\left(\theta_k\right) = \sqrt{\frac{2}{N}} \cos\left(m\theta_k + \phi\left(\theta_k\right)\right), \quad u_m^s\left(\theta_k\right) = \sqrt{\frac{2}{N}} \sin\left(m\theta_k + \phi\left(\theta_k\right)\right), \tag{2.35}$$

$$v_n^c\left(\theta_k\right) = \sqrt{\frac{2}{N}} \cos\left(n\theta_k\right), \quad u_n^s\left(\theta_k\right) = \sqrt{\frac{2}{N}} \sin\left(n\theta_k\right). \tag{2.36}$$

We expand the equation (2.32)

$$\begin{aligned} \tilde{\Omega}_{mn} &= \frac{1}{N}\left[\omega(0) + (-1)^{m-n}\omega(\pi) + 2\sum_{0<\theta_k<\pi} \omega\left(\theta_k\right) \cos\left(\theta_k(m-n) + \phi\left(\theta_k\right)\right)\right] \\ &= \frac{\omega(0)}{N} + (-1)^{m-n}\frac{\omega(\pi)}{N} + \sum_{0<\theta_k\leq\pi} \omega\left(\theta_k\right)\left(u_m^c\left(\theta_k\right) v_n^c\left(\theta_k\right) + u_m^s\left(\theta_k\right) v_n^s\left(\theta_k\right)\right), \end{aligned} \tag{2.37}$$

Now imposing $u^s(0) = v^s(\pi) = 0$y$u^c(0) = v^c(\pi) = \frac{1}{\sqrt{N}}$, to rewrite $\tilde{\Omega}_{m,n}$

$$\tilde{\Omega}_{mn} = \sum \omega\left(\theta_k\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right) + u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right). \tag{2.38}$$

Therefore, the upper-right block of (2.29) of the Hamiltonian reads

$$H = \sum_{m,n=0}^{N-1}\frac{i}{4}\sum_{\theta_k=0}^{\pi}\omega\left(\theta_k\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right) + u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right)\left[\hat{\gamma}_n, \hat{\gamma}_{m+N}\right], \tag{2.39}$$

and rearranging things, we get

$$H = \sum_{\theta_k=0}^{\pi}\omega\left(\theta_k\right)(\underbrace{\left[\hat{\gamma}_k^c, \hat{\gamma}_{k+N}^c\right]}_{1-2\sigma_k^z} + \underbrace{\left[\hat{\gamma}_k^s, \hat{\gamma}_{k+N}^s\right]}_{1-2\sigma_k^z}), \tag{2.40}$$

where we have used

$$\hat{\gamma}_k^{c,s} := \sum_n u_n^{c,s}\left(\theta_k\right)\hat{\gamma}_n, \quad \hat{\gamma}_{k+N}^{c,s} := \sum_n v_n^{c,s}\left(\theta_k\right)\hat{\gamma}_{n+N}. \tag{2.41}$$

Now recall that the Fermionic covariance matrix is defined by (2.10), then, the transformation that brings Ω into its Williamson form, does the same on the FMC. Thus the upper-right block of the FMC is position space is

$$\begin{aligned}
\tilde{\Gamma}_{mn} &= \sum_{\theta_k}^{\pi}\left[m^c\left(\theta_k\right)u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right) + m^s\left(\theta_k\right)u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right] \\
&= \sum_{\theta_k}^{\pi}\left(\frac{m^c\left(\theta_k\right) + m^s\left(\theta_k\right)}{2}\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right) + u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right) \\
&+ \sum_{\theta_k}^{\pi}\left(\frac{m^c\left(\theta_k\right) - m^s\left(\theta_k\right)}{2}\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right) - u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right),
\end{aligned} \tag{2.42}$$

where $m^{c,s}(\theta_k) = n^{c,s}(\theta_k) - \frac{1}{2}$, being $n^{c,s}(\theta_k$ the "cosine" ("sine") fermion occupation number of the mode labeled by k.
Now let $m^{\pm}\left(\theta_k\right) = \frac{m^c(\theta_k)\pm m^s(\theta_k)}{2}$. We can undo the transformation from (2.32) to (2.38) to have

$$\tilde{\Gamma}_{mn} = \overbrace{\sum_{\theta_k}^{\pi} m^+\left(\theta_k\right)e^{i\phi(\theta_k)}e^{i(n-m)\theta_k}}^{\tilde{\Gamma}_{mn}^+} + \underbrace{\sum_{\theta_k}^{\pi} m^-\left(\theta_k\right)e^{i\phi(\theta_k)}e^{i(n+m)\theta_k}}_{\tilde{\Gamma}_{mn}^-}. \tag{2.43}$$

We notice that $\tilde{\Gamma}_{mn}^+$ is circulant, whereas $\tilde{\Gamma}_{mn}^-$ is not, However, observe that $\tilde{\Gamma}_{mn}^+ = \tilde{\Gamma}_{mn'}^-$, with n' a change on the index $n \to -n'$. As the figure 2.1 shows, this transformation can be interpreted as a rotation over the circle.

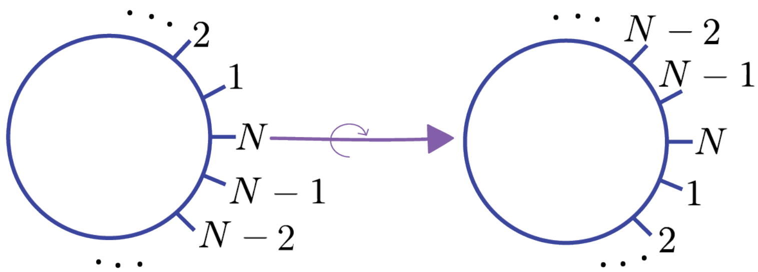

Figure 2.1: Meaning of the relabel done in the circulant matrix, which can be seen as a reflection over the circle.

Explicitly we can write that if $\tilde{\Gamma}^{+}_{mn}$ has the shape

$$\begin{pmatrix} a_0 & a_{-1} & \cdots & a_2 & a_1 \\ a_1 & a_0 & \cdots & a_3 & a_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{-1} & a_{-2} & a_{-3} & \cdots & a_0 \end{pmatrix}, \tag{2.44}$$

then $\tilde{\Gamma}^{-}_{mn}$ will be given by

$$\begin{pmatrix} a_0 & a_1 & \cdots & a_{-2} & a_{-1} \\ a_1 & a_2 & \cdots & a_{-1} & a_0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{-1} & a_0 & a_1 & \cdots & a_{-2} \end{pmatrix}, \tag{2.45}$$

which we name anticirculant.

In this way, both the circulant and the anticirculant parts of Γ are computed as Fourier transforms of the vectors $m^{+}(\theta_k)e^{i\phi(\theta_k)}$ and $m^{-}(\theta_k)e^{i\phi(\theta_k)}$ respectively.

Here we spot 3 things. First, the FCM always can be written as a circulant matrix Γ^{+} plus an anticirculant matrix Γ^{-}. Second, in the ground state, the FCM is circulant because the fermion occupation numbers $n^c(\theta_k) = n^s(\theta_k) = 0, \forall k$. Third, for a generic excited state, we have that in average the FCM matrix is always circulant, because $\langle n^c(\theta_k)\rangle = \langle n^s(\theta_k)\rangle$.

2.4.6. Local modes in the fermionic chains

For our purpose, it is important to study the behaviour of reduced states in the chain, that is, we want to look into small portions of size L in a translationally invariant chain of size N ($N > L$). To do this it is convenient to write the modes of the small chain of size L in terms of the modes of whole chain of size N.

Since we work with a one-dimensional, translationally invariant closed chain of N free fermions, with local interactions, it is often useful to expand the annihilation/creation operators $\hat{a}_x$ ($\hat{a}_l^\dagger$) at the site x ($x = 0, 1, \ldots, N-1$) in terms of their counterpart in the plane-wave basis, obtained through

$$\hat{b}_q = \frac{e^{i\eta_q}}{\sqrt{N}} \sum_{x=0}^{N-1} e^{-i\theta_q x} \hat{a}_x, \tag{2.46}$$

and its hermitian conjugate, with $\theta_q = \frac{2\pi}{N} q$, and η_q is a phase to be adjusted. By doing so, we showed that in the case of the XY model, the Hamiltonian (2.18) became diagonal on the operators $\hat{b}_q$, explicitly we showed that the Hamiltonian had the form

$$\hat{H} = \sum_{q=0}^{N} \Lambda(\theta_q) \hat{b}_q^\dagger \hat{b}_q, \tag{2.47}$$

with $\Lambda(\theta_q)$ given by (2.24) in the XY model.

We consider now a set of local plane wave modes for the portion of the chain of length L comprising the sites $x = 0, 1, \ldots, L-1$,

$$\tilde{b}_k = (-1)^k \frac{e^{-i\tilde{\theta}_k/2}}{\sqrt{L}} \sum_{x=0}^{L-1} e^{-i\tilde{\theta}_k x} \hat{a}_x, \tag{2.48}$$

where similarly as in the case of the large chain of size N, we take $\tilde{\theta}_k = \frac{2\pi}{L} k$. By introducing the modes of the local chain in this way, the plain wave modes will diagonalize any Hamiltonian of the form (2.47), with N replaced by L.

We now expand the local operators $\tilde{b}_k$ in terms of the global operators $\hat{b}_q$ that are defined over the chain of size N. The relation between the two becomes

$$\tilde{b}_k = \frac{1}{\sqrt{NL}} \sum_{q=0}^{N-1} D_L(\theta_q - \tilde{\theta}_k) \hat{b}_q, \tag{2.49}$$

where we chose appropriately $\eta_q = \theta_q (L-1)/2$, and

$$D_L(\theta) = \frac{\sin\left(\frac{\theta}{2} L\right)}{\sin\left(\frac{\theta}{2}\right)}, \tag{2.50}$$

is the Dirichlet kernel [65].

Now consider an excited state $|\vec{n}\rangle$ of the chain, described by a set of excitation numbers $\vec{n} \equiv (n_0, n_1, \ldots, n_{N-1})$, where $n_q \in \{0, 1\}$. In particular the state satisfies

$$\langle \vec{n}| \hat{b}_q^\dagger \hat{b}_{q'} |\vec{n}\rangle = \delta_{q,q'} n_q, \quad \langle \vec{n}| \hat{b}_q \hat{b}_{q'} |\vec{n}\rangle = \langle \vec{n}| \hat{b}_q^\dagger \hat{b}_{q'}^\dagger |\vec{n}\rangle = 0, \tag{2.51}$$

for all q, q'. To any excited state, we can associate the correspondent $L \times L$ FCM, that is

$$A_{kk'}(\vec{n}) = \langle \vec{n}| \tilde{b}_k^\dagger \tilde{b}_{k'} |\vec{n}\rangle = \frac{1}{NL} \sum_{q=0}^{N-1} D_L(\theta_q - \tilde{\theta}_k) D_L(\theta_q - \tilde{\theta}_{k'}) n_q. \tag{2.52}$$

Therefore, we have a full characterization of any subchain in the state $|\vec{n}\rangle$, or equivalently in the state

$$\rho_L(\vec{n}) = \mathrm{Tr}_{N-L} |\vec{n}\rangle \langle \vec{n}| , \tag{2.53}$$

the corresponding partial density matrix on the subchain.

2.5. Error correcting code theory

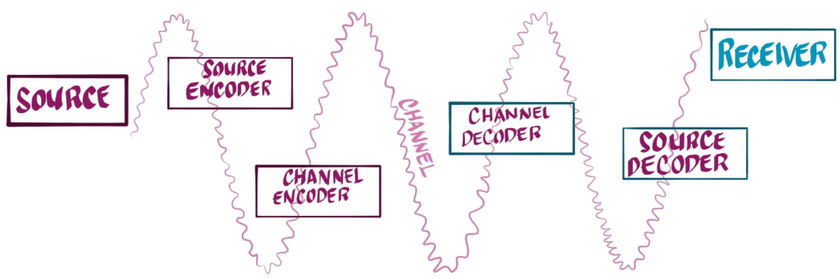

Figure 2.2: Representation of the scheme of communication.The encoding system introduces some redundancy into the transmitted vector $\mathbf{x}$. The decoding system uses this known redundancy to deduce from the received vector $\mathbf{y}$ both the original source vector and the noise introduced by the channel.

We will now turn to study error correcting codes. We start by motivating the problem of error correction codes as a mechanism to understand communication, then we move to introduce basic definitions to formalise the problem of error correcting codes, and afterwards, we will move to study the case of random minimum distance codes and some interesting results about them.

Every day we communicate over noisy channels such as in telephone lines, over which two devices communicate digital information through a bunch of cables. When we think in designing these channels we have as main purpose to be able to transmit information in a reliable way while dealing with errors induced by the noise in the channel. Information theory and coding theory offer a way to study communications as C. Shannon pointed out in 1948 [66]. In order to stablish communication, there are some main ingredients that C. Shannon explicitly presented in his paper [66].

As we illustrate in figure 2.2 in order to pass a message through a channel to then someone receive it, we add encoders before the channel and decoders after it. The encoders encode the source message **x** into a transmitted message **y**, adding redundancy to the original message in some way. The channel adds noise to the transmitted message, yielding a received message **y**. The decoders use the known redundancy introduced by the encoding system to infer both the original signal **s** and the added noise.

2.5.1. Channel coding

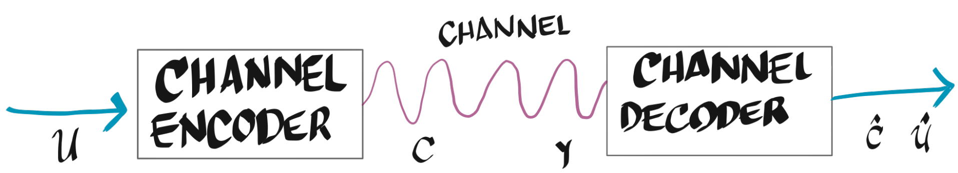

Figure 2.3: Channel coding

For the purpose of our work, the model of channel we will use, is the *discrete probability channel*: a probabilistic channel S is defined as a triple (F, Φ, Prob), where F is a finite *input alphabet*, Φ is a finite *output alphabet*, and Prob is a conditional probability distribution

$$\text{Prob}\{\mathbf{y} \text{ received } |\mathbf{x} \text{ transmited}\}, \tag{2.54}$$

defined for every pair $(\mathbf{x}, \mathbf{y}) \in F^m \times \Phi^m$, where m ranges over all positive integers and F^m/Φ^m denotes the set of all words of length m over F/Φ. It is important to clarify that we assume that the channel neither deletes nor inserts symbols; that is, the length of the output word **y** always equals the length of the input word **x**.

The input channel encoder is an *information word* **u** out of M possible information words (see figure 2.3). The channel encoder generates a *codeword* $\mathbf{c} \in F^n$ that is input to the channel. The resulting output of the channel is a received word $\mathbf{y} \in \Phi^n$, which is fed into the channel decoder. the decoder, in turn, produces a *decoded codeword* $\hat{\mathbf{c}}$ and a *decoded information word* $\hat{\mathbf{u}}$, with the aim of having $\mathbf{c} = \hat{\mathbf{c}}$ and $\mathbf{u} = \hat{\mathbf{u}}$. this implies that the channel encoder needs to be such that the mapping $\mathbf{u} \mapsto \mathbf{c}$ is one to one.

Definition 2.5.1 (Rate). *The rate of the channel encoder is defined as*

$$R = \frac{\log_{|F|} M}{n}. \tag{2.55}$$

If all information words have the same length over F, then this length is given by the numerator, $\log_{|F|} M$, in the expression for R. Since the mapping of the encoder is one-to-one, we have that $R \leq 1$.
In the case where the input alphabet F has the same size as the output alphabet Φ, it will be convenient to assume that $F = \Phi$ and that the elements of F form a finite Abelian group. We then say that the channel is an *additive channel*.

Given an additive channel, let $\mathbf{x}$ and $\mathbf{y}$ be input and output words, respectively, both in F^m. the *error word* is defined as the difference $\mathbf{x} - \mathbf{y}$, where the subtraction is taken component by component. The action of the channel can be described as adding an error word $\mathbf{e} \in F^m$ to the input word $\mathbf{x}$ to produce the output word $\mathbf{y} = \mathbf{x} + \mathbf{e}$.

When F is an Abelian group, it contains the zero (or unit) element. The *error locations* are indexes of the nonzero entries in the error word $\mathbf{e}$. Those entries are referred to as the *error values*.

2.5.2. Binary symmetric channel (BSC)

As an illustrative example, consider the input and output alphabets are $F = \Phi = \{0, 1\}$, and for every two binary words $\mathbf{x}$, $\mathbf{y}$ for a given length m,

$$\begin{aligned} &\text{Prob}\{\mathbf{y} \text{ received } |\mathbf{x} \text{ transmited}\} \\ &\qquad = \prod_{j=1}^{m} \text{Prob}\{y_j \text{ received } |x_j \text{ transmited}\}, \end{aligned} \tag{2.56}$$

where for every $x, y \in F$,

$$\text{Prob}\{y \text{ received } |x \text{ transmited}\} = \begin{cases} 1-p & \text{if } y = x \\ p & \text{if } y \neq x \end{cases} \tag{2.57}$$

The parameter p is a real number $0 \leq p \leq 1$ adn is called the *crossover probability* of the channel.

the action of the BSC can be described as flipping each input bit with probability p, independently of the past or the future. The channel is called symmetric since the probability of the flip is the same regardless of weather the input is 0 or 1.

2.5.3. Block codes

An (n, M) (*block*) code over a finite alphabet F is a nonepty subset $\mathfrak{C}$ of size M of F^n. The parameter n is called the *code length* and M is the *code size*. The *information length* or *codeword length* of $\mathfrak{C}$ is defined by $k = \log_{|F|} M$, and the *rate* is $R = k/n$. The range of the mapping defined by the channel encoder in figure 2.3 forms an (n,M) code, and this in which the term (n, M) code will be used. The elements of a code are called *codewords*.

In addition to the length and size of a code, we will be interested in quantifying how much the codewords in the code differ from each other. To do this, we will introduce the following definitions.

Definition 2.5.2 (Hamming distance)**.** *Let F be an alphabet. The Hamming distance between two codewords $\mathbf{x}, \mathbf{y} \in F^n$ is the number of coordinates on which $\mathbf{x}$ and $\mathbf{y}$ differ. the Hamming distance will be denoted by $d(\mathbf{x}, \mathbf{y})$.*

It is easy to verify that the Hamming distance satisfies the following properties of a metric for every three words $\mathbf{x}, \mathbf{y}, \mathbf{z} \in F^n$.

(I) $d(\mathbf{x}, \mathbf{y}) \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{y}$).

(II) Symmetry: $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.

(III) The triangle inequality: $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$.

Definition 2.5.3 (Hamming weight)**.** *Let F be an Abelian group. The Hamming weight of $\mathbf{e} \in F^n$ is the number of nonzero entries in $\mathbf{e}$. We denote the Hamming weight by $w(\mathbf{e})$.*

Notice that for every two words $\mathbf{x}, \mathbf{y} \in F^n$,

$$d(\mathbf{x}, \mathbf{y}) = w(\mathbf{y} - \mathbf{x}). \tag{2.58}$$

Turning back now to block codes, let $\mathfrak{C}$ be an (n, M) code over F with $M > 1$. The *minimum distance* of $\mathfrak{C}$ is the minimum Hamming distance between any two distinct codewords of $\mathfrak{C}$;

that is, the minimum distance d is given by

$$d = \min_{\mathbf{c}_1, \mathbf{c}_2 \in \mathfrak{C} : \mathbf{c}_1 \neq \mathbf{c}_2} d(\mathbf{c}_1, \mathbf{c}_2). \tag{2.59}$$

An (n, M) with minimum distance d is often called (n, M, d) code. We will sometimes use the notation $d(\mathfrak{C})$ for the minimum distance of a given code C.

2.5.4. Decoding

Let $\mathfrak{C}$ be an (n, M, d) code over an alphabet F and let S be the channel defined by the triple (F, Φ, Prob). A decoder for the code $\mathfrak{C}$with respect to the channel S is a function

$$\mathcal{D} : \Phi^n \to \mathfrak{C}. \tag{2.60}$$

The *decoding error probability* P_{err} of $\mathcal{D}$ is defined by

$$P_{\mathrm{err}} = \max_{\mathbf{c} \in \mathfrak{C}} P_{\mathrm{err}}(\mathbf{c}), \tag{2.61}$$

where

$$P_{\mathrm{err}}(\mathbf{c}) = \sum_{\mathbf{y} : \mathcal{D}(\mathbf{y}) \neq \mathbf{c}} \mathrm{Prob}\{\mathbf{y} \text{ received } | \mathbf{c} \text{ transmited}\}. \tag{2.62}$$

Note that $P_{\mathrm{err}}(\mathbf{c}$ is the probability that the code word $\mathbf{c}$ will be decoded erroneously, given that $\mathbf{c}$ was transmitted.

Maximum-likelihood decoding

We next consider particular strategies for codes and channels. Given an (n, M, d) code $\mathfrak{C}$ over F and a channel $S = (F, \Phi, \mathrm{Prob})$, a *maximum-likelihood decoder* (MLD) for $\mathfrak{C}$ with respect to S is the function $\mathcal{D}_{MLD} : \Phi^n \to \mathfrak{C}$ defined as follows: for every $\mathbf{y} \in \Phi^n$, the value $\mathcal{D}_{MLD}$ equals the codeword $\mathbf{c} \in \mathfrak{C}$ that maximizes the probability

$$\mathrm{Prob}\{\mathbf{y} \text{ received } | \mathbf{c} \text{ transmited}\}, \tag{2.63}$$

In the case of a tie between two (or more) codewords, we choose one of the tying codewords arbitrarily. Hence, $\mathcal{D}_{MLD}$ is well-defined for the code $\mathfrak{C}$ and the channel S.

Decoder in repetition code

As an alternative to the MLD we will consider a *majority vote decoder*, we consider $\mathfrak{C}$ to be the binary $(3, 2, 3)$ *repetition code* and let S be the BSC with crossover probability p.

The binary $(3, 2, 3)$ repetition code is the code $\{000, 111\}$ over $F = \{0, 1\}$, which has information length $\log_2 2 = 1$ and rate $1/3$.
Define a decoder $\mathcal{D} : \{0, 1\}^3 \to \mathfrak{C}$ as follows:

$$\mathcal{D}(000) = \mathcal{D}(001) = \mathcal{D}(010) = \mathcal{D}(100) = 000, \tag{2.64}$$

and

$$\mathcal{D}(011) = \mathcal{D}(101) = \mathcal{D}(110) = \mathcal{D}(111) = 111. \tag{2.65}$$

This decoder is known as the vote majority decoder.

The probability P_{err} equals the probability of having two or more errors

$$\begin{aligned} P_{\text{err}} = P_{\text{err}}(000) = P_{\text{err}}(111) &= \binom{3}{2} p^2(1-p) + \binom{3}{3} p^3 \\ &= 3p^2 - 3p^3 + p^3 \\ &= 3p^2 - 2p^3. \end{aligned} \tag{2.66}$$

Notice then that the error probability is dominated by the probability that two bits in a block of tree are flipped, which scales as p^2. Since our goal is to have decoders with small P_{err}, from equation (2.66) we conclude that P_{err} is smaller than p when $p < 1/2$, which simply means that coding has improved the probability of error per message, compared to uncoded transmission. However, the price is reflected in the rate: three bits are transmitted for every bit implies a rate of $\log_2 M/n = 1/3$.

Nearest-codeword decoder

We next consider a particular decoding strategy. A *nearest-codeword decoder* for an (n, M) code $\mathfrak{C}$ over F is a function $F^n \to \mathfrak{C}$ whose value for every word $\mathbf{y} \in F^n$ is a closes codeword in $\mathfrak{C}$ to $\mathbf{y}$, where the term "closest" is with respect to the Hamming distance. A nearest-codeword decoder for $\mathfrak{C}$ is a decoder for $\mathfrak{C}$ with respect to any additive channel whose input and output alphabets are F.

Capacity of the binary symmetric channel

So far we have seen that coding allows to reduce the decoding error probability P_{err}, at the expense of transmitting at lower rates. As we will discuss, it is possible to achieve arbitrarily small values of P_{err}, while still transmitting at rates that are bounded away from 0.

Define the *binary entropy* function $\mathfrak{H} : [0,1] \to [0,1]$ by

$$\mathfrak{H}(x) = -x \log_2 x - (1-x)\log_2(1-x), \tag{2.67}$$

where $\mathfrak{H}(0) = \mathfrak{H}(1) = 0$.

Now let S be the BSC with crossover probability p. The *capacity* of S is given by

$$\mathrm{Cap}(S) = 1 - \mathfrak{H}(p). \tag{2.68}$$

These definitions will allows us to present a special case of two fundamental results in information theory. These results state the the capacity of a channel is the largest rate at which information can be transmitted reliably through that channel

Theorem 2.5.1 (Shannon coding theorem for the BSC)**.** *Let s be the BSC with crossover probability p and let R be a real in the range $0 \leq R \leq \mathrm{Cap}(S)$. there exist an infinite sequence of (n_i, M_i) block codes over $F\{0,1\}$, $i = 1,2,\ldots$, such that $(\log_2 M_i)/n_i \geq R$ and, for maximum-likelihood decoding for those codes (with respect to S), the decoding error probability P_{err} approaches 0 as $i \to \infty$.*

Theorem 2.5.2 (Shannon converse coding theorem for the BSC)**.** *Let S be the BSC channel with crossover probability p and let R be a real greater than $\mathrm{Cap}(S)$. Consider any infinite sequence of (n_i, M_i) block codes over $F = \{0,1\}, i = 1,2,\ldots$, such that $(\log_2 M_i)/n_i \geq R$ and $n_1 < n_2 < \cdots < n_i < \cdots$. Then, for any decoding scheme for those codes (with respect to S), the decoding error probability P_{err} approaches 1 as $i \to \infty$.*

These theorems are know as the noisy-channel theorems for the special case of the BSC. A problem with announcing the theorems in this particular way is that are quite general results; meaning that the theorems only says that reliable communication with crossover probability p and rate R can be achieved by using code blocks with sufficiently large code length n_i[2]. The theorem does not say how large n needs to be to achieve given values of R and p. Particularly, it is possible to show that P_{err} in these theorems can be guaranteed to decrease exponentially with the code length n_i.

For a discrete memoryless channel, a length code n and rate R, there exist a block code (n, M) whose average probability of error satisfies:

$$P_{\mathrm{err}} \leq \exp(-nE_r(R)), \tag{2.69}$$

[2]To see further details about the proof see [67]

where $E_r(R)$ is known as the *random-coding exponent* of the channel, a convex, decreasing, positive function of R for $0 \leq R \leq \mathrm{Cap}(S)$. the random-coding exponent is also known as the reliability function [68,69]. $E_r(R)$ approaches zero as $R \to \mathrm{Cap}(S)$. As we will see the computation of the random-coding exponent is in general a challenging task on which much effort has been expended.

2.5.5. Levels of error handling

While the setting in previous sections was probabilistic, we will move to identify error words that are generated by an additive channel and are always recoverable, as long as the transmitted codewords are taken from a block code whose minimum distance is sufficiently large. The results we present here are combinatorial, meaning that they do not depend on the particular conditional probability of the channel.

Error correction

We consider channels $S = (F, \Phi, \mathrm{Prob})$ with $\Phi = F$.

Given an (n, M, d) code $\mathfrak{C}$ over F, let $\mathbf{c} \in \mathfrak{C}$ be the transmitted codeword and $\mathbf{y} \in F^n$ be the received word. By *error* we mean the event of changing an entry in the codeword $\mathbf{c}$. The number of errors equals $d(\mathbf{y}, \mathbf{c})$, and the error locations are indexes of the entries in which $\mathbf{c}$ and $\mathbf{y}$ differ.

The task of error correction is recovering the error locations and the error values. In the next lemma we show that errors are always recoverable, as long as their number does not exceed a certain threshold, which depends on the code $\mathfrak{C}$.

Lemma 2.5.3. *Let $\mathfrak{C}$ be an (n, M, d) code over F. There is a decoder $\mathcal{D} : F^n \to \mathfrak{C}$ that recovers correctly every pattern up to $\lfloor (d-1)/2 \rfloor$ errors for every channel $S = (F, F, \mathrm{Prob})$.*

Proof. Let $\mathcal{D}$ be a nearest-codeword decoder, namely, $\mathcal{D}(\mathbf{y})$ is a closes (with respect to the Hamming distance) codeword in $\mathfrak{C}$ to $\mathbf{y}$. Let $\mathbf{c}$ and $\mathbf{y}$ be the transmitted codeword and the received word, respectively, where $d(\mathbf{y}, \mathbf{c}) \leq (d-1)/2$. Suppose to the contrary that $\mathbf{c}' = \mathcal{D}(\mathbf{y}) \neq \mathbf{c}$. By the way $\mathcal{D}$ is defined,

$$d(\mathbf{y}, \mathbf{c}') \leq d(\mathbf{y}, \mathbf{c}) \leq (d-1)/2. \tag{2.70}$$

So, by the triangle inequality,

$$d \leq d(\mathbf{c}, \mathbf{c}) \leq d(\mathbf{y}, \mathbf{c}) + d(\mathbf{y}, \mathbf{c}') \leq d-1, \tag{2.71}$$

which is a contradiction. $\square$

2.5.6. Random minimum distance codes and its error exponents

Having in mind that for minimum distance codes it is possible to perfectly recover the message up to a number of error of $\lfloor (d-1)/2 \rfloor$, we will study the case of random minimum distance codes over the memoryless binary channel with crossover probability $p = 1/2$.

Let $\mathfrak{C}$ be a random block code $(n, M \equiv 2^{nR})$ over $F = \{0,1\}$ and rate R. This is known as a random code ensemble (RCE) and is obtained by uniformly random choosing each of the n bits in each of the M codewords independently. Equivalently, each of the 2^{nM} possible binary codes of length n and rate R in the RCE is assigned probability 2^{-nM}.

Since these codes are randomly chosen we can ask ourselves if these codes will have minimum distance and if so we will be interested in the distribution of distances. To answer this, we first compute the probability that a given random codeword $\mathbf{x}_i$ of length n will be at a Hamming distance $d = \delta n$ from an arbitrary binary n-tuple $\mathbf{y}$, where $\delta = d/n$ is known as the *relative distance*. This probability is independent of $\mathbf{y}$ and equals [24]

$$\mathrm{Prob}(d(\mathbf{x}_i, \mathbf{y}) = d) = \binom{n}{d} \left(\frac{1}{2}\right)^d \left(\frac{1}{2}\right)^{n-d} \equiv 2^{-n(1-\mathfrak{H}(\delta))}, \tag{2.72}$$

where $\mathfrak{H}(\delta)$ is the binary entropy defined in (2.67). Under this RCE, two Hamming distances $d(\mathbf{x}_i, \mathbf{x}_j)$, $d(\mathbf{x}'_i, \mathbf{x}'_j)$ are independent random variables unless $\{i,j\} = \{i', j'\}$ or $\{i,j\} = \{j', i'\}$.

Consider the number of unordered pairs of codewords $(\mathbf{x}_i, \mathbf{x}_j)$ with $i \neq j$ in $\mathfrak{C}$ at Hamming distance d apart

$$\mathcal{S}_{\mathfrak{C}}(d) = \sum_{i=0}^{M-1} \sum_{j=0}^{i-1} \Theta\{d(\mathbf{x}_i, \mathbf{x}_j) = d\}, \tag{2.73}$$

where $\Theta\{d(\mathbf{x}_i, \mathbf{x}_j) = d\}$ is the indicator of the event in the brackets, that is, $\Theta\{d(\mathbf{x}_i, \mathbf{x}_j) = d\}$ is equal to 1 if $d(\mathbf{x}_i, \mathbf{x}_j) = d$ and to 0 otherwise. Then $\mathcal{S}_{\mathfrak{C}}(d)$ is a sum of $\binom{M}{2}$ pairwise-independent, identically distributed random variables $\Theta\{d(\mathbf{x}_i, \mathbf{x}_j) = d\}$, each with mean

$$\mathbb{E}\Theta = \mathrm{Prob}\{d(\mathbf{x}_i, \mathbf{x}_j) = d\}, \tag{2.74}$$

where $\mathbb{E}$ refers to the expected value or simply the average of a random variable which is obtained by

$$\mathbb{E}(X) = \sum_{x \in X} \mathrm{Prob}(X = x) x, \tag{2.75}$$

with X a random variable.

The variance of $\Theta\{d(\mathbf{x}_i, \mathbf{x}_j) = d\}$ is

$$\mathrm{Var}[\Theta] = \mathbb{E}\Theta^2 - (\mathbb{E}\Theta)^2 = \mathbb{E}\Theta - (\mathbb{E}\Theta)^2 < \mathbb{E}\Theta, \tag{2.76}$$

where we have used the fact that $\Theta^2 = \Theta$ since Θ is a $\{0,1\}$-valued function. We have then that in the case of the BSC, $\mathcal{S}_{\mathfrak{C}}(d)$ is a sum of $\binom{M}{2}$ independent and identically distributed random variables.

$$\mathbb{E}\mathcal{S}_{\mathfrak{C}}(d) = \binom{M}{2}\mathbb{E}\Theta = 2^{n(2R-1+\mathfrak{H}(\delta))}, \tag{2.77}$$

where we have used the fact that $M \sim 2^{nR}$, $\binom{M}{2} = M(M-1)/2 \sim M^2 = 2^{2nR}$. Since the variance of a sum of uncorrelated random variables is equal to the sum of their variances

$$\mathrm{Var}(\mathcal{S}_{\mathfrak{C}}(d)) = \binom{M}{2}\mathrm{Var}\,\Theta < \mathbb{E}\mathcal{S}_{\mathfrak{C}}(d). \tag{2.78}$$

As we mentioned before, we will be interested in study the minimum distance $d(\mathfrak{C})$ that this particular kind of codes have. To do this we first announce the next three lemmas that will let us introduce one of the main results in [24] about the distribution of minimum distance in random codes.

Lemma 2.5.4 (Markov inequality)**.** *Let X be a non-negative real random variable, and $\alpha > 0$, then the probability that X is at least α is at most the expectation of X divided by α,*

$$\mathrm{Prob}(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha} \tag{2.79}$$

Lemma 2.5.5 (Chebyshev's inequality)**.** *Let X be real random variable with finite expected value $\mathbb{E}(X) = \mu$ and finite variance $\mathrm{Var}(X) = \sigma^2$, then for all $k > 0$,*

$$\mathrm{Prob}(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2} \tag{2.80}$$

Proof. Apply Markov's inequality with $\alpha \equiv k^2$ and take the random variable $X - \mu$

$$\mathrm{Prob}\left((X-\mu)^2 \geq k^2\right) \leq \frac{\mathbb{E}(X-\sigma)}{k^2} = \frac{\sigma^2}{k^2} \tag{2.81}$$

□

Lemma 2.5.6 (Chernoff bound)**.** *Let X be real random variable with finite mean, for all $\alpha > 0$ and $t > 0$,*

$$\mathrm{Prob}(X \geq \alpha) \leq \frac{\mathbb{E}(e^{tX})}{e^{t\alpha}} \tag{2.82}$$

Proof. First, take the exponential at both sides of the expression $\mathrm{Prob}(X \geq \alpha)$,

$$\mathrm{Prob}(X \geq \alpha) = \mathrm{Prob}(e^{tX} \geq e^{t\alpha}), \tag{2.83}$$

then we apply Markov's inequality and conclude that

$$\mathrm{Prob}(X \geq \alpha) = \mathrm{Prob}(e^{tX} \geq e^{t\alpha}) \leq \frac{\mathbb{E}(e^{tX})}{e^{t\alpha}} \tag{2.84}$$

□

In particular, we can optimize over t to get

$$\mathrm{Prob}(X \geq \alpha) \leq \min_{t>0} \frac{\mathbb{E}(e^{tX})}{e^{t\alpha}}. \tag{2.85}$$

Now we are ready to state the the following theorem [24].

Theorem 2.5.7 (Distance distribution in RCE). *For $0 \leq R < 1/2$ and any $\varepsilon > 0$, the probability that a code length n and rate R from the RCE has relative minimum distance less than $\delta_{GV}(2R) - \varepsilon$, with $\delta_{GV}(R)$ the solution to the equation $\mathfrak{H}(\delta) = 1 - R$, goes to zero exponentially as $n \to \infty$. For $0 \leq R < 1$, if $d = n\delta$ is such that*

$$\delta_{GV}(2R) + \varepsilon \leq \delta \leq 1 - \delta GV(2R) - \varepsilon, \tag{2.86}$$

then the probability that the number of codeword pairs at a distance d satisfies $\mathcal{S}_{\mathfrak{C}}(d) \doteq 2^{n(2R-1+\mathfrak{H}(\delta))}$ goes to one as $n \to \infty$.

Proof. For a given value of the code value of the code rate R we can choose d such that $d/n \to \delta \leq \delta_{\mathrm{GV}}(2R) - \varepsilon$. Then

$$\mathrm{Prob}\left\{\mathcal{S}_{\mathfrak{C}}(d) \geq 1\right\} \leq \mathbb{E}\mathcal{S}_{\mathfrak{C}}(d) \doteq 2^{-n(1-\mathfrak{H}(\delta)-2R)} \to 0, \tag{2.87}$$

which in other words it tells us that with probability differing from 1 by an exponentially falling quantity, there will be no pairs at distance d. Conversely, if $\delta_{\mathrm{GV}}(2R) + \varepsilon < \delta < 1 - \delta_{\mathrm{GV}}(2R) - \varepsilon$, then $1 - \mathcal{H}(\delta) < 2R$ and the average of number of pairs $\mathbb{E}\mathcal{S}_{\mathfrak{C}}(d))$ at a distance d is exponentially large. To see this, we can use the Chebyshev's inequality, so for any $\delta > 0$, we have

$$\mathrm{Prob}\left\{|\mathcal{S}_{\mathfrak{C}}(d) - \mathbb{E}\mathcal{S}_{\mathfrak{C}}(d)| \geq \begin{pmatrix} M \\ 2 \end{pmatrix} \alpha\right\} \leq \frac{\mathbb{E}\Theta}{\begin{pmatrix} M \\ 2 \end{pmatrix} \alpha^2}, \tag{2.88}$$

by choosing $\alpha \doteq 2^{-n(1-\mathfrak{H}(\delta)+\Delta)} < \mathbb{E}\Theta$ for any $\Delta > 0$, then we have

$$\mathrm{Prob}\left\{|\mathcal{S}_{\mathfrak{C}}(d) - \mathbb{E}\mathcal{S}_{\mathfrak{C}}(d)| > \begin{pmatrix} M \\ 2 \end{pmatrix} \alpha\right\} \leq \frac{2\mathbb{E}\Theta}{M(M-1)\alpha^2} \doteq 2^{-n(2R-1+\mathfrak{H}(\delta)-2\Delta)}. \tag{2.89}$$

The exponent on the right-hand side can be made positive by choosing Δ small enough. This establishes the fact that $\mathcal{S}_{\mathfrak{C}}(d) \doteq 2^{n(2R-1+\mathcal{H}(\delta))}$ for the chosen value of d with probability tending to one as $n \to \infty$. $\square$

Alternative this result can be expressed as follows. With probability $1-2^{-\zeta(n)}$ as $n \to \infty$, the relative minimum distance of a code drawn from the RCE will be approximately $\delta GV_(2R)$ for $0 \leq R \leq \frac{1}{2}$ and 0 for $\frac{1}{2} \leq R \leq 1$.

Define the distance distribution the random code $\mathfrak{C}$ as

$$\mathcal{N}_{\mathfrak{C}}(d) = \frac{2}{M}\mathcal{S}_{\mathfrak{C}}(d), \quad (d = 0, 1, \ldots, n). \tag{2.90}$$

Thus $\mathcal{N}_{\mathfrak{C}}(d)$ is the average over the M codewords $\mathbf{x}_i$ of the number of other codewords $\mathbf{x}_j, j \neq i$, at Hamming distance d from $\mathbf{x}_i$. We therefore have shown that the complete average distance distribution over RCE is

$$\mathcal{N}_{RCE}(d) = \frac{2}{M}\mathbb{E}\mathcal{S}_{\mathfrak{C}}(d) \doteq 2^{n(R-1+\mathfrak{H}(\delta))}. \tag{2.91}$$

Since the probability in (2.89) tends to zero exponentially and since there are only $N+1$ different values of distance d, theorem 2.5.7 implies that for almost all codes in RCE $\mathcal{N}_{\mathfrak{C}}(d) \doteq \mathcal{N}_{RCE}(d)$ for all d such that $\delta_{\mathrm{GV}}(2R) + \varepsilon \leq \delta \leq 1 - \delta_{\mathrm{GV}}(2R) - \varepsilon$. However for $\delta \leq \delta_{\mathrm{GV}}(2R) - \varepsilon$ or $\delta \geq 1 - \delta_{\mathrm{GV}}(2R) + \varepsilon$, the $\mathcal{N}_{RCE}(d) = 0$.

In the next chapter, we will develop the correspondent connection between systems whose Hamiltonians are quadratic in annihilation/creation fermionic operators, and code theory. Particularly, we will use the last results in the distribution distance of random codes to explicitly show how is possible to treat fermionic excited states as random codes that have a definite energy value. A similar result as the one in theorem 2.5.7 will be shown to argue that there are exponentially large Hilbert subspaces that fulfil exactly the condition of ultra-orthogonality.

Chapter 3

Studying ultra-orthogonality in fermionic systems

In the last two chapters we have discussed the foundations of statistical mechanics and we provided the needed theoretical background to delimit our problem to the case of a special kind of fermionic systems. Specifically, in the first chapter we discussed the foundations of statistical mechanics and we introduced typicality as an individualist approach of thermalisation, then we motivated the property of ultra-orthogonality as a consequence of typicality and we showed that ultra-orthogonality is related to equilibration as an instantaneous phenomenon. Afterwards, in the second chapter, we introduced all the theoretical background to delimit our problem to a special kind of fermionic systems, finally in the last section, we talked about coding theory and we showed an important result in random codes. This chapter will be devoted to connect all the concepts presented in previous chapters and show an explicit connection between random code theory and fermionic systems. Moreover, we will show how this connection allows us to show that for a specific kind of fermionic systems, it is possible to find exponentially large Hilbert subspaces such that ultra-orthogonality can be guaranteed.

3.1. Delimiting our problem

In order to delimit our problem we have first to recall some concepts that were introduced previously. First, in the first part of the second chapter, we introduced the fermionic quadratic Hamiltonians in the operators of creation/annihilation. As we mentioned in that moment, these quadratic Hamiltonians have the interesting property that not only the ground state but every eigenstate representing a certain number of excitations of quasi-particles, belong to the so-called class of fermionic Gaussian states. We also discussed that this particular class of states are completely characterised by its second order correlations as the Wick theorem states. This characterisation motivated us to study these kind of systems in terms of the Majorana fermions $\{\hat{\gamma}_i\}$. The algebra generated by the operators $\{\hat{\gamma}_i\}$ is known as the Clifford algebra, which for the case of systems of N modes we denoted by $\mathcal{C}_{2N}$. Since the canonical anticommutation relation is given by $\{\hat{\gamma}_i, \hat{\gamma}_j\} = 2\delta_{ij}$, an arbitrary operator $\hat{X} \in \mathcal{C}_{2N}$ can be represented as a polynomial in $\{\hat{\gamma}_i\}$ [52], namely

$$\hat{X} = \alpha\hat{\mathbb{I}} + \sum_{p=1}^{N} \sum_{1\leq a_1<\cdots<a_p\leq N} \alpha_{a_1,\ldots,a_p} \hat{\gamma}_{a_1}\hat{\gamma}_{a_1+N}\cdots\hat{\gamma}_{a_p}\hat{\gamma}_{a_p+N}, \tag{3.1}$$

where $\alpha = 2^{-N}\,\mathrm{Tr}(\hat{X})$ and $\alpha_{a_1,\ldots,a_p}$ are the expansion coefficients of X. The operator $\hat{X} \in \mathcal{C}_{2N}$ is called even (odd) if it involves only even (odd) powers of the generators $\hat{\gamma}$. Notice that in the expansion of $\hat{X}$ the operators $\hat{\gamma}_i$ and $\hat{\gamma}_{i+N}$ are involved, operators that are defined in (2.6) through the operators of creation and annihilation at the site i. Hence, we can denote by $x_i \in \{0, 1\}$ the presence or absence of the operator $\hat{\gamma}_i$. Thus, we can build the binary vector $\vec{x}$ that will give information about the presence or absence of the operator $\hat{\gamma}_i$ in the expansion of $\hat{X}$. Similarly we can construct the vector $\vec{y}$ which will tells us the information about the presence or absence of the operator $\hat{\gamma}_{i+N}$. Therefore, we have that the expansion of $\hat{X} \in \mathcal{C}_{2N}$, given in (3.1), can alternatively be written as

$$\hat{X} = \sum_{\vec{x},\vec{y}} f(\vec{x},\vec{y})\vec{\gamma}(\vec{x},\vec{y}), \tag{3.2}$$

where

$$\vec{\gamma}(\vec{x},\vec{y}) = \hat{\gamma}_1^{x_1}\hat{\gamma}_{1+N}^{y_1}\cdots\hat{\gamma}_N^{x_N}\hat{\gamma}_{2N}^{y_N}, \tag{3.3}$$

where x_i (y_i) is 1 when the operator $\hat{\gamma}_i$ ($\hat{\gamma}_{i+N}$) appears and 0 otherwise, and $f(\vec{x},\vec{y})$ is the coefficient expansion of $\hat{X}$. By using this notation, we are able to write products of the form $\vec{\gamma}(\vec{x},\vec{y})\vec{\gamma}(\vec{x}',\vec{y}')$ as

$$\vec{\gamma}(\vec{x},\vec{y})\vec{\gamma}(\vec{x}',\vec{y}') = e^{i\phi(\vec{x},\vec{y},\vec{x}',\vec{y}')}\vec{\gamma}(\vec{x}+\vec{x}',\vec{y}+\vec{y}'), \tag{3.4}$$

where the term $e^{i\phi(\vec{x},\vec{y},\vec{x}',\vec{y}')}$ refers to the phase that appears when dealing with products of Majorana operators, and that in general depends on the sequences $\vec{x},\vec{y},\vec{x}'$ and $\vec{y}'$ (strictly speaking, the phase depends on the involved products of Majorana operators that may appear), and where the sum $\vec{x}+\vec{x}'$ is taken $\mod 2$, since it is not possible to have quadratic terms in any Majorana operator [52].

On the other hand, in the first chapter we saw that ultra-orthogonality appeared as a consequence of typicality, thus ultra-orthogonality has to be studied in the same framework of typicality. It means that the Hilbert subspace in which the states of the universe are sampled has to be constrained to a global restriction (e.g. a defined constant energy E). To guarantee that a global restriction is imposed over the states we study, we will make use of a very well-known technique, named Gibbs sampling, for obtaining sequences of observations which are approximated from a specified multivariate probability distribution [70–72]. But before we go any further, it is important to translate the concepts of universe, environment and subsystem, defined in chapter one, to the fermionic system of N modes. For the purpose of this work we will identify the universe as the system of N fermionic modes and the subsystem as the portion that corresponds to L fermionic modes, with $L \ll N$.

Now consider an excited state $|\vec{n}\rangle$ of the fermionic system, described by the Hamiltonian given in equation (2.1). The excited state is described by the set of excitation numbers $\vec{n} \equiv (n_1, n_2, \ldots, n_N)$, where $n_q \in \{0,1\}$. Our goal is then to independently sample the occupations n_q accordingly to the Fermi-Dirac distribution, that is:

$$\mathrm{Prob}(\vec{n} \mid \beta) = \prod_{q=1}^{N} \mathrm{Prob}\left(n_q \mid \beta\right), \quad \mathrm{Prob}\left(n_q \mid \beta\right) = \frac{e^{-\beta\omega(\theta_q)n_q}}{\sum_{n_q} e^{-\beta\omega(\theta_q)n_q}}, \tag{3.5}$$

where $\theta_q = \frac{2\pi}{N}q$, $\omega(\theta_q)$ is the spectrum of the correspondent of the system described by a Hamiltonian of the form (2.1) (where in the special case of the XY model, is given by the equation (2.24)), and β refers to the inverse temperature. This way of sampling reproduce a well-defined value of the energy when $N \to \infty$, condition that is important to us since the energy will play the role of the global constraint for our system. To see the reason that the energy will be well-defined when $N \to \infty$, we first calculate the average number of excitation in the mode at angle θ_q

$$v\left(\theta_q \mid \beta\right) \equiv \langle n_q \rangle_\beta = \frac{1}{e^{\beta\omega(\theta_q)} + 1}, \tag{3.6}$$

while the variance in the number of excitations is given by

$$\left\langle n_q^2 \right\rangle_\beta - \langle n_q \rangle_\beta^2 = \frac{1}{e^{\beta\omega(\theta_q)} + 1} = v\left(\theta_q\right)\left(1 - v\left(\theta_q\right)\right). \tag{3.7}$$

Thus the mean energy is

$$\langle E\rangle_\beta = \sum_q v_q\left(\theta_q \mid \beta\right)\omega_q\left(\theta_q\right) = N \oint_N \frac{d\theta}{2\pi} f(\theta \mid \beta)\omega(\theta), \tag{3.8}$$

where $\oint_N$ denotes the Riemman sum approximation to the corresponding integral with N subdivisions. As we are interested in the thermodynamic limit $N \to \infty$, in most cases one may replace the integral for $\oint_N$. Similarly, the energy variance of the sampled states is given by

$$\left\langle \Delta E^2\right\rangle_\beta = \sum_q v\left(\theta_q\right)\left(1 - v\left(\theta_q\right)\right)\omega_q^2\left(\theta_q\right) = N \oint_N \frac{d\theta}{2\pi} v\left(\theta_q\right)\left(1 - v\left(\theta_q\right)\right)\epsilon^2(\theta). \tag{3.9}$$

Thus, Gibbs sampling provides an even sampling of states within ΔE of the energy $\langle E\rangle$, where $\Delta E/\langle E\rangle \sim O(N^{-1/2})$.

Summarising,we have shown that an operator $\hat{X} \in \mathcal{C}_{2N}$ can be expanded in terms of two sequences $\vec{x}$ and $\vec{y}$ that represent the presence or the absence of the operator $\hat{\gamma}_i$ and $\hat{\gamma}_{i+N}$ respectively. Moreover, we showed that excited states of the fermionic system of N modes fulfilling a condition about the energy, can be sampled independently from the Fermi-Dirac distribution. Having these ideas in mind, we are ready to show our first result.

3.2. Studying ultra-orthogonality

In order to study ultra-orthogonality for the case of systems that can be mapped to quadratic fermionic Hamiltonians, we will consider two randomly chosen, excited states $|\vec{n}_1\rangle$ and $|\vec{n}_2\rangle$. As mentioned in the last part of the chapter one, ultra-orthogonality is a property that whenever we take the partial trace over the environment of the exterior product of two states of the universe ($|\vec{n}_1\rangle$ and $|\vec{n}_2\rangle$) is approximately zero, that is $\mathrm{Tr}_{\mathcal{E}} |\vec{n}_1\rangle\langle\vec{n}_2| \approx 0$, thus we will be interested in the states of the form:

$$\hat{\rho}_{12} = |\vec{n}_1\rangle\langle\vec{n}_2|\,. \tag{3.10}$$

Thus by taking the partial trace over the $N - L$ sites of the system we wet

$$\hat{X}_{12} = \mathrm{Tr}_{N-L}\,\hat{\rho}_{12} = \mathrm{Tr}_{N-L}\,|\vec{n}_1\rangle\langle\vec{n}_2|\,. \tag{3.11}$$

Notice that since we took the respective partial trace over the $N - L$ modes, the operator $\hat{X}_{12}$ is an operator that can be expanded in terms of majorana operators of L modes, that is, $\hat{X}_{12} \in \mathcal{C}_{2L}$ and therefore has an expansion in terms of the Majorana operators $\hat{\gamma}_i$ and $\hat{\gamma}_{i+L}$, with $i = 1, 2, \ldots L$. We then write $\hat{X}_{12}$ with the notation in (3.2) where we will keep the sequences $\vec{x}$ and $\vec{y}$ to be of length N, but for this case we will put zeros in the $N - L$ remaining spaces, thus making explicit the absence of the $N - L$ remaining operators. Thus $\hat{X}_{12}$ will be written as

$$\hat{X}_{12} = \sum_{\vec{x},\vec{y}} f(\vec{x},\vec{y})\vec{\gamma}(\vec{x},\vec{y}), \tag{3.12}$$

and the sequences $\vec{x}, \vec{y}$ are sequences of length N that have zeros in the $N - L$ positions.

Thus, our task will be to find the coefficient $f(\vec{x},\vec{y})$ to know the expansion of $\hat{X}_{12}$. In order to do this, we multiply equation by $\vec{\gamma}^\dagger(\vec{x}',\vec{y}')$ and take the partial trace over L

$$\mathrm{Tr}_L\left(\hat{X}_{12}\vec{\gamma}^\dagger(\vec{x}',\vec{y}')\right) = \sum_{\vec{x},\vec{y}} f(\vec{x},\vec{y})\,\mathrm{Tr}_L\left(\vec{\gamma}(\vec{x},\vec{y})\gamma^\dagger(\vec{x}',\vec{y}')\right). \tag{3.13}$$

Using the equation (3.4) in (3.13), the right hand side of the equation will become $\delta_{\vec{x}+\vec{x}'}\delta_{\vec{y}+\vec{y}'}$. Thus the coefficient $f(\vec{x},\vec{y})$ is given by

$$f(\vec{x}',\vec{y}') = \frac{1}{2^L}\,\mathrm{Tr_L}\left(\hat{X}_{12}\gamma^\dagger(\vec{x}',\vec{y}')\right) = \frac{1}{2^L}\left\langle \vec{n}_2\right|\gamma^\dagger(\vec{x}',\vec{y}')\left|\vec{n}_1\right\rangle. \tag{3.14}$$

So in order to find the coefficient $f(\vec{x},\vec{y})$ we have to explicitly show how the operator $\vec{\gamma}(\vec{x},\vec{y})$ acts over $|\vec{n}_1\rangle$ ($|\vec{n}_2\rangle$). To explicitly show this, we recall from equation (2.11) that there is an orthogonal transformation that brings the Majorana operators acting on the spatial modes of fermionic states to the normal modes, that is

$$\overbrace{\gamma_{i_1}\gamma_{i_2}\ldots\gamma_{i_k}}^{\text{Spacial modes}} = O_{i_1\alpha_1}O_{i_2\alpha_2}\ldots O_{i_k\alpha_k}\underbrace{\gamma_{\alpha_1}\gamma_{\alpha_2}\ldots\gamma_{\alpha_k}}_{\text{Normal modes}}. \tag{3.15}$$

where the O_{i_j,α_j} refers to the orthogonal transformation discussed in section 2.3. Now, notice that the operators $\vec{\gamma}$ obtained after this orthogonal transformation will be diagonal over the sequences $\vec{x}$, $\vec{y}$. This implies that the are two possible ways of getting an excited state $|\vec{n}_1\rangle$ through the function $\vec{\gamma}$ that is written on the normal mode basis. The first one is by applying the the operator $\vec{\gamma}(\vec{n}_1, 0)$ over the vacuum state $|0\rangle$, that is

$$|n_1\rangle = \vec{\gamma}(\vec{n}_1, 0)\,|0\rangle\,, \tag{3.16}$$

and the second is

$$|n_1\rangle = \vec{\gamma}(0, \vec{n}_1)\,|0\rangle\, e^{i\phi(\vec{n_1})}. \tag{3.17}$$

So, the $\vec{x}$ sequence takes the vacuum state $|0\rangle$ and turns it into an excited state by changing the zeros to ones, and alternatively the $\vec{y}$ sequence transforms the vacuum state $|0\rangle$ to an excited state and adds a phase.

Therefore, the right hand-side of equation (3.14) yields

$$\langle\vec{n}_2|\,\vec{\gamma}(\vec{x},\vec{y})\,|\vec{n}_1\rangle \propto \delta_{\vec{n}_1+\vec{x}+\vec{y},\vec{n}_2}e^{i\phi(\vec{n}_1,\vec{n}_2,\vec{x},\vec{y})} = \delta_{\vec{x}+\vec{y},\vec{n}_2+\vec{n}_1}e^{i\phi(\vec{n}_1,\vec{n}_2,\vec{x},\vec{y})}, \tag{3.18}$$

where the last equality is guaranteed since the sums are taken to be modulo 2.

The coefficient $f(\vec{x},\vec{y})$, is then given by

$$f(\vec{x},\vec{y}) = \frac{1}{2^L}\sum_{\vec{x}',\vec{y}'}\mathcal{U}_{\vec{x}\vec{x}'}\mathcal{V}_{\vec{y}\vec{y}'}\underbrace{\langle\vec{n}_2|\,\vec{\gamma}(\vec{x},\vec{y})\,|\vec{n}_1\rangle}_{\propto\delta_{\vec{n}_1+\vec{n}_2,\vec{x}+\vec{y}}}. \tag{3.19}$$

where $\mathcal{U}_{\vec{x}\vec{x}'}$ ($\mathcal{V}_{\vec{y}\vec{y}'}$) refers to the orthogonal matrix transformation that brings the operators γ_i (γ_{i+L}) to its normal mode representation. Note that the vector $\vec{n}_1+\vec{n}_2$ is the vector that has a one in the $i-$th component when both vectors ($\vec{n}_1$ and $\vec{n}_2$) differ in the position i, that is, $n_{1_i}\neq n_{2_i}$. Thus, the vector $\vec{n}_1+\vec{n}_2$ is interpreted as the error vector. of the sequences

The result in equation (3.19) is extremely important because it shows that when we work with states like $\hat{X}_{12}$, if the error vector $\vec{n}_1+\vec{n}_2$ has a larger number of ones than the vector $\vec{x}+\vec{y}$, the coefficient $f(\vec{x},\vec{y})$ will immediately become zero. Moreover, notice that we have defined $\vec{x}$ and $\vec{y}$ to be sequences that allows us to expand the operator $\hat{X}_{12}\in\mathcal{C}_{2L}$, meaning that the vector $\vec{x}+\vec{y}$ can have at most L ones. Therefore, we have that the coefficient $f(\vec{x},\vec{y})$ will be equal to zero when the number of ones in the vector $\vec{n}_1+\vec{n}_2$ exceeds L.

This particular result brings questions like, "how likely is it to have more than L error when $N\gg L$?" and "what happens when we have less errors than L?". In next section we will be addressing the first question, namely, we will interpret the result found in equation (3.19) in terms of a random minimum distance code and we will show that the largest Hilbert subspace in which ultra-orthogonality holds is exponentially large.

3.3. Fermionic random minimum codes.

In order to make an explicit connection between minimum distance codes, described in the last part of the second chapter, and the found result in the previous section, consider the *error vector* $\vec{e}_{12}$ to be the sum (modulo 2) of the sequences $\vec{n}_1$ and $\vec{n}_2$

$$\vec{e}_{12} = \vec{n}_1+\vec{n}_2. \tag{3.20}$$

As we pointed out before, the vector $\vec{e}_{12}$ has ones in the entries where $\vec{n}_1$ and $\vec{n}_2$ differ, that is

$$e_{12_i} = \begin{cases} 0 & n_{1_i} = n_{2_i}, \\ 1 & n_{1_i} \neq n_{2_i}. \end{cases} \tag{3.21}$$

We denote by d the Hamming distance between the sequences $\vec{n}_1$ and $\vec{n}_2$, or alternatively by the Hamming weight of $\vec{e}_{12}$, that is

$$d = d(\vec{n}_1, \vec{n}_2) = w(\vec{e}_{12}). \tag{3.22}$$

In the previous section showed that there is a specific distance $d > L$ at which the operator $\hat{X}_{12}$, given by equation (3.12), is equal to zero. Therefore, we can write the problem of working with states $|\vec{x}^{(i)}\rangle$ in terms of a (N, M, d) code $\mathfrak{C}$ over the alphabet $F = \{0, 1\}$ and the channel S given by the memoryless binary symmetric channel. Thus, the code $\mathfrak{C}$ is a binary code with code length N and code size M. Denoting the information length of the code $\mathfrak{C}$ as $k = \log_2 M$, and the rate $R = k/N$, we have that the code $\mathfrak{C}$ will be form by the codewords

$$\mathfrak{C} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(2^k)}\} \quad \mathbf{x}^{(i)} \in \{0, 1\}^N. \tag{3.23}$$

thus if we assign each of the codewords $\mathbf{x}^{(i)}$ to the respective state $|\vec{x}^{(i)}\rangle$ such that the states $\{|\vec{x}^{(1)}\rangle, |\vec{x}^{(2)}\rangle, \ldots\}$ are mutually ultra-orthogonal, we can ask ourselves about the Hilbert subspace $\mathcal{H}_{\mathfrak{C}}$ spanned by the states $|\vec{x}^{(i)}\rangle$, that is

$$\mathcal{H}_{\mathfrak{C}} = \text{Span}\left(|\vec{x}^{(1)}\rangle, |\vec{x}^{(2)}\rangle, \ldots, |\vec{x}^{(2^k)}\rangle\right). \tag{3.24}$$

Notice that by assigning the codewords $\mathbf{x}^{(i)}$ to the respective state $|\vec{x}^{(i)}\rangle$, the condition of ultra-orthogonality is guaranteed since the code $\mathcal{C}$ is a minimum distance code (N, M, d) with $d > L$, and as we pointed out in chapter two, a minimum distance code has the property that the minimum Hamming distance of any two distinct codewords of $\mathfrak{C}$ is d

$$d = \min_{\mathbf{x}^{(i)}, \mathbf{x}^{(j)} \in \mathfrak{C} : \mathbf{x}^{(i)} \neq \mathbf{x}^{(j)}} d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}). \tag{3.25}$$

Now take an arbitrary state $|\psi(t)\rangle \in \mathcal{H}_{\mathfrak{C}}$. We can expand it in terms of the elements of its basis as

$$|\psi(t)\rangle = \sum_{\vec{n} \in \mathfrak{C}} \psi(\vec{n}, t) |\vec{n}\rangle. \tag{3.26}$$

Defining $\hat{\rho} = |\psi(t)\rangle \langle\psi(t)|$ the state of the system with N modes, we have that by construction of the Hilbert subspace $\mathcal{H}_{\mathfrak{C}}$, when we take the partial trace over the $N - L$ fermionic modes, we will end up with a time-independent state, that is

$$\rho_L = \mathrm{Tr}_{N-L}\left(|\psi(t)\rangle\langle\psi(t)|\right) = \sum_{\vec{n}\in\mathfrak{C}} |\psi(\vec{n})|^2\, \mathrm{Tr}_{N-L}\, |\vec{n}\rangle\langle\vec{n}|\,, \tag{3.27}$$

where we have ignored the crossed terms

$$\sum_{\vec{n}\neq\vec{m}} \psi(\vec{n},t)\psi(\vec{m},t)^*\, \mathrm{Tr}_{N-L}\, |\vec{n}\rangle\langle\vec{m}|\,. \tag{3.28}$$

since by construction they are all zero ($\mathrm{Tr}_{N-L}\, |\vec{n}\rangle\langle\vec{m}| = 0$).

Thus, every state $|\psi(t)\rangle \in \mathcal{H}_{\mathfrak{C}}$ of the system of N fermionic modes has the property that its reduce state is instantaneously stationary, condition that we can write as

$$\frac{d\rho_L}{dt} = 0. \tag{3.29}$$

Having this in mind, we can ask ourselves about the largest minimum distance code (N, M, d) that can be built to alternatively answer what is the largest Hilbert subspace where ultra-orthogonality holds.

In order to do this we will follow thoroughly the steps of section 2.5.6 to compute the average number of codeword that are at a given distance d. Note that the probability that two codewords differ in d sites was computed in the last part of the second chapter. However, the probability was computed by supposing that each entry had equal probability to occur, and this is not the case for excited fermionic states, since as we saw in equation (3.5), each entry of the sequence has a different probability to occur and it depends on the mode q. Therefore, the probability that two sequences differ in d sites has to be modified for the case of excited states.

Note that the probability that at position q the entry is given by the Fermi-Dirac distribution. Thus the probability $p(\theta_q)$ that two sequences differ in position q will be given by

$$p(\theta_q) = \binom{2}{1} \mathrm{Prob}(n_q|\beta)\left(1 - \mathrm{Prob}(n_q|\beta)\right), \tag{3.30}$$

where $\mathrm{Prob}(n_q|\beta)$ corresponds to the Fermi-Dirac distribution in equation (3.5).

If we define the random variable $X_i(\theta_q)$ to take the value 1 with probability $p(\theta_q)$ and 0 with probability $1 - p(\theta_q)$. Explicitly that is

$$\mathrm{Prob}(X_i(\theta_q) = k) = \begin{cases} p(\theta_q) & k = 1, \\ 1 - p(\theta_q) & k = 0. \end{cases} \tag{3.31}$$

Define the sum of the random variables $X = \sum_k X_i(\theta_q)$. Notice that X will be nothing but a random variable that equals the total number of errors between two sequences.

We will be interested to see what is the probability that the random variable X is equal or larger than d, that is

$$\mathrm{Prob}(X \geq d), \tag{3.32}$$

To compute this probability is in general a complicated task and involve complex expression; instead, we will bound the probability in equation (3.32). To do this we will make use of the independence of the random variables $X_i(\theta_q)$, and a well-know bound defined in equation (2.82), named *Chernoff bound*.

Let S be a real random number grater than zero ($S > 0$), then from equation (2.83) we know that

$$\mathrm{Prob}(X \geq d) = \mathrm{Prob}(e^{SX} \geq e^{Sd}) \leq \left\langle e^{SX} \right\rangle e^{-Sd}, \tag{3.33}$$

where $\left\langle e^{SX} \right\rangle$ refers to the expected value.

Particularly we are interested in the optimal S, then, as we pointed out in equation (2.85), this optimization can be done by taking the minimum S in inequality (3.33)

$$\mathrm{Prob}(e^{SX} \geq e^{Sd}) \leq \min_{S>0} \left\langle e^{SX} \right\rangle e^{-Sd}. \tag{3.34}$$

We then have to compute the expected value $\left\langle e^{SX} \right\rangle$ in order to proceed with the optimization. Since X is the sum of random independent variables we have that

$$\left\langle e^{SX} \right\rangle = \prod_q E(e^{SX(\theta_q)}) = \prod_q \left(1 + p(\theta_k)(e^S - 1)\right) = e^{\sum_q \log(1+p(\theta_q)(e^S-1))}. \tag{3.35}$$

Thus, putting equation (3.34) and (3.35), we get

$$\mathrm{Prob}(e^{SX} \geq e^{Sd}) \leq \min_{S>0} e^{\sum_q \log(1+p(\theta_q)(e^S-1)-Sd)}. \tag{3.36}$$

Defining $r(\delta)$ as the minimization exponent in equation (3.36), we get

$$r(\delta) = \min_S \frac{1}{N} \left(\sum_q \log(1 + p(\theta_q)(e^S - 1) - S\delta \right), \tag{3.37}$$

where $\delta = d/N$. We are interested in the case $N \to \infty$, meaning that $r(\delta)$ can be written as

$$r(\delta) = \min_S \oint_N \frac{d\theta}{2\pi} \log\left(1 + p(\theta)(e^S - 1)\right) - S\delta, \tag{3.38}$$

where $\oint_N$ is the one defined in equation (3.8). Note that $r(\delta)$ corresponds to the random-coding exponent associated with the fermionic random states, which we discussed in equation

(2.69). For these reason we refer to the binary sequence of excitations present in a fermionic state as a *fermionic random code*.

Even though it is not possible to analytically solve equation (3.38) for S, we can differentiate it with respect to S to get a transcendental equation

$$\delta = \oint_N \frac{d\theta}{2\pi} \frac{p(\theta)e^S}{1 - p(\theta) + p(\theta)e^S}. \tag{3.39}$$

By using equation (3.38) we have that after the optimization over S, the inequality in (3.33) takes its minimum value [24]. Therefore, if similarly as in section 2.5.6, we ask ourselves about the number of unordered pairs of codewords $(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ in $\mathfrak{C}$ with $i \neq j$, at a Hamming distance d apart ($\mathcal{S}_\mathcal{C}$). As we showed in equation (2.77), the average number of unordered pairs of codewords at a distance d apart is given by

$$\langle S_\mathfrak{C}(d) \rangle = \binom{M}{2} 2^{-Nr(\delta)}. \tag{3.40}$$

Remembering that $k = \log_2 M$ and $R = k/N$, we can replace $M = 2^{NR}$ in equation (3.40) and conclude

$$\langle S_\mathcal{C}(d) \rangle = 2^{N(2R - r(\delta))}. \tag{3.41}$$

From this expression we conclude that whenever we work with rates higher than $r(\delta)/2$, the average number of unordered pairs of codewords at a distance d is exponentially large with the number of modes N in the fermionic system. And conversely, whenever we are at lower rates than $r(\delta)/2$, this number goes to zero exponentially.

With the result shown in equation (3.41) we conclude that when the rate of the random fermionic code is larger than rates higher than $r(\delta)/2$, there is an associated exponentially large Hilbert subspace $\mathcal{H}_\mathfrak{C}$ in which every state taken from it has the property that its correspondent reduced state is automatically stationary, in other words, we are able to find exponentially large Hilbert subspaces in which the property of ultra-orthogonality holds exactly.

Chapter 4

Conclusions and Perspectives

Throughout all this document, we have presented a series of arguments to show, that ultra-orthogonality is intimately related with the mechanism of equilibration as an instantaneous phenomenon. As we mentioned, ultra-orthogonality is connected to thermalisation problem, since the overwhelming majority of states in Hilbert space, are such that its correspondent reduced states approximately coincide with the thermal state, hence, the state of equilibrium we expect to have for most cases should be the canonical state. Particularly, we addressed the problem when Ultra Orthogonality holds exactly, that is, the equilibrium state is automatically reached when computing the correspondent reduced state of the Universe.

This work was devoted to show that for the a special kind of Fermionic systems, there exists a particular sort of pure dynamical states, such that when we take the correspondent partial trace over its environment, the resultant reduced states are automatically in its correspondent state of equilibrium when the condition of having more differences than the number of modes in the subsystem of L modes is fulfilled. In our results an interpretation in terms of binary sequences of the excited states was needed to make our conclusions.

Not satisfied with finding this result we decided to estimate the size of the correspondent Hilbert subspace associated with the states that fulfil exactly the property of ultra-orthogonality, in the sense that all reduced states, will be automatically stationary states (constant states over time). By using the formalism of random-coding exponents, explained in chapter two, it was possible to provide an expression to the large deviation present in the Fermionic systems. Namely, the exponent provided an estimate of the expected number of random Fermionic minimum codes that fulfil ultra-orthogonality. Particularly we showed that whenever this exponent is larger than zero, it is possible to show that there is exponentially large Hilbert subspaces fulfilling the condition of ultra-orthogonality. In other words, the main result of our work is that it is possible to build exponentially large Hilbert subspaces in which any fully interacting state will be immediately stationary when we compute the partial trace over the correspondent $N - L$ modes.

The result in this work brought more questions than answers to us, where one of the first questions was "what happens if think this problem in terms of linear codes in which a restriction over the energy is imposed?". This particular question came to our mind because as it is shown in the work of A. Brag. et. al. [24] linear random codes have a better exponent of error than the one obtained from the random codes. From our point of view, we consider that this could guide to interesting insights about equilibration in these kind of systems. However, this particular problem was not addressed since liner superposition of codewords may end up not conserving the restriction of energy, therefore we consider that another study will be needed to confirm if this can be done and provide the respective insights to equilibration.

We want to point out that the physical meaning of these quantities was not fully understood and the question "what exactly is the physical meaning of these exponents?" remained open for us. What we want to stress here is that indeed, there was an attempt to find some physical interpretation about these quantities but not conclusive result was found.

Finally, the case when the number of differences between two random fermionic codewords is less than L remain an open problem. Nonetheless, we strongly belief that this case has to lead to similar conclusion about ultra-orthogonality, that is, we expect that when the number of differences d is less than the number of modes in the subsystem L, ultra-orthogonality has also to appear.

Bibliography

[1] S. Popescu, A. J. Short, and A. Winter. The foundations of statistical mechanics from entanglement: Individual states vs. averages.

[2] S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghì. Canonical typicality. 96(5):050403.

[3] J. Gemmer and M. Michel. Thermalization of quantum systems by finite baths. 73(1):1–7.

[4] S. Popescu, A. J. Short, and A. Winter. Entanglement and the foundations of statistical mechanics. 2(11):754–758.

[5] S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zanghi. Approach to thermal equilibrium of macroscopic quantum systems. *Phys. Rev. E*, 81:011109, Jan 2010.

[6] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner. Quantum thermalization through entanglement in an isolated many-body system.

[7] C. Gogolin and J. Eisert. Equilibration, thermalisation, and the emergence of statistical mechanics in closed quantum systems. 79(5):056001.

[8] Shang-Keng M. *Statistical Mechanics*. Wspc, illustrated edition edition.

[9] I. E. Farquhar. *Ergodic Theory in Statistical Mechanics*. John Wiley & Sons Inc, 1st ed. edition edition.

[10] R. Jancel. *Foundations of Classical and Quantum Statistical Mechanics: International Series of Monographs in Natural Philosophy*. Pergamon.

[11] M. Horodecki, J. Oppenheim, and A. Winter. Partial quantum information. 436(7051):673–676.

[12] J. M. Deutsch. Thermodynamic entropy of a many-body energy eigenstate. 12(7):075021.

[13] N. Singh. On the foundations of statistical mechanics: a brief review. 27(6):1330003. Publisher: World Scientific Publishing Co.

[14] J. Gemmer, M. Michel, and G. Mahler. *Quantum Thermodynamics: Emergence of Thermodynamic Behavior Within Composite Quantum Systems*, volume 657 of *Lecture Notes in Physics*. Springer Berlin Heidelberg.

[15] V. D. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds*. Springer.

[16] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society.

[17] H. Tasaki. From quantum dynamics to the canonical distribution: General picture and a rigorous example. 80(7):1373–1376.

[18] N. Linden, S. Popescu, A. J. Short, and A. Winter. Quantum mechanical evolution towards thermal equilibrium. 79(6):061103.

[19] N. Linden, S. Popescu, A. J. Short, and A. Winter. On the speed of fluctuations around thermodynamic equilibrium. 12(5):055021.

[20] A. S. L. Malabarba, L. P. García-Pintos, N. Linden, T. C. Farrelly, and A. J. Short. Quantum systems equilibrate rapidly for most observables. 90(1):012121.

[21] M. Srednicki. Chaos and quantum thermalization. 50(2):888–901.

[22] J. M. Deutsch. Quantum statistical mechanics in a closed system. 43:2046–2049.

[23] M. Rigol and M. Srednicki. Alternatives to eigenstate thermalization. 108(11).

[24] A. Barg and G.D. Forney. Random codes: minimum distances and error exponents. 48(9):2568–2573. Conference Name: IEEE Transactions on Information Theory.

[25] J. W. Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. C. Scribner's Sons.

[26] L. Boltzmann. *Zur Priorität der Auffindung der Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und dem Principe der kleinsten Wirkung*, volume 219. Wiley Online Library, 1871.

[27] L.R. Statistical physics: (1962 brandeis lectures) vol. 3 (w. a. benjamin, new york, 1963. vii—252 p. $ 5.45 paperbound, $ 8.80 clothbound). 48:697.

[28] P. Ehrenfest and T. Ehrenfest. *The conceptual foundations of the statistical approach in mechanics, by Paul and Tatiana Ehrenfest. Translated by Michael J. Moravcsik*. Ithaca, N.Y., Cornell University Press [1959].

[29] P. L. Garrido, S. Goldstein, and J. L. Lebowitz. The boltzmann entropy for dense fluids not in local equilibrium. 92(5):050602.

[30] S. Goldstein and Joel L. Lebowitz. On the (boltzmann) entropy of nonequilibrium systems. 193(1):53–66.

[31] L. Boltzmann. *Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie:(vorgelegt in der Sitzung am 8. Februar 1866)*. Staatsdruckerei, 1866.

[32] L. Boltzmann. *Über die Beziehung zwischen dem zweiten Hauptsatze des mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht*. Kk Hof-und Staatsdruckerei, 1877.

[33] S. Goldstein. Boltzmann's approach to statistical mechanics. In Jean Bricmont, Giancarlo Ghirardi, Detlef Dürr, Francesco Petruccione, Maria Carla Galavotti, and Nino Zanghi, editors, *Chance in Physics: Foundations and Perspectives*, Lecture Notes in Physics, pages 39–54. Springer.

[34] Sheldon Goldstein, Joel L. Lebowitz, Roderich Tumulka, and Nino Zanghi. Gibbs and boltzmann entropy in classical and quantum mechanics.

[35] E. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

[36] E. T. Jaynes. Gibbs vs boltzmann entropies. *American Journal of Physics*, 33(5):391–398, 1965.

[37] J. Butterfield. On symplectic reduction in classical mechanics. In J. Butterfield and John Earman, editors, *Philosophy of Physics*, Handbook of the Philosophy of Science, pages 1 – 131. North-Holland. ISSN: 18789846.

[38] G. Belot. The representation of time and change in mechanics. In Jeremy Butterfield and John Earman, editors, *Philosophy of Physics*, Handbook of the Philosophy of Science, pages 133 – 227. North-Holland. ISSN: 18789846.

[39] E. T. Jaynes. Information theory and statistical mechanics. ii. *Phys. Rev.*, 108:171–190, Oct 1957.

[40] J. Von Neumann. *Mathematical foundations of quantum mechanics.* Princeton, N.J. : Princeton University Press.

[41] W. H. McCrea. The principles of statistical mechanics. by r. c. tolman. pp. xix, 661. 40s. 1938. international series of monographs on physics. (oxford). *The Mathematical Gazette*, 23(256):415–417, 1939.

[42] D. ter Haar. Chapter 6 - the ensembles in quantum statistics. In D. ter Haar, editor, *Elements of Statistical Mechanics (Third edition)*, pages 216 – 267. Butterworth-Heinemann, Oxford, third edition edition, 1995.

[43] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. 45(4):1216–1227. Conference Name: IEEE Transactions on Information Theory.

[44] Giancarlo C., G. T. Landi, and S. Eliëns. On the strong subadditivity of the r\'enyi entropies for bosonic and fermionic gaussian states. 99(4):045155.

[45] R. J. Glauber. Density operators for fermions. In *Quantum Theory of Optical Coherence*, pages 595–631. John Wiley & Sons, Ltd. Section: 16 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527610075.ch16.

[46] Y. Niu, S. B. Chung, C. Hsu, I. Mandal, S. Raghu, and S. Chakravarty. Majorana zero modes in a quantum ising chain with longer-ranged interactions. 85(3):035110.

[47] A. F. Reyes-Lega. Some aspects of operator algebras in quantum physics. pages 1–74.

[48] M. Chung and I. Peschel. Density-matrix spectra of solvable fermionic systems. 64(6):064412.

[49] M. Leijnse and K. Flensberg. Introduction to topological superconductivity and majorana fermions. 27(12):124003.

[50] L. G. Molinari. Notes on wick's theorem in many-body theory.

[51] A. Botero and B. Reznik. BCS-like modewise entanglement of fermion gaussian states. 331(1):39 – 44.

[52] S. Bravyi. Lagrangian representation for fermionic linear optics.

[53] E. Lieb, T. Schultz, and D. Mattis. Two soluble models of an antiferromagnetic chain. 16(3):407–466.

[54] J. I. Latorre, E. Rico, and G. Vidal. Ground state entanglement in quantum spin chains. 4(1):48–92.

[55] S. Katsura. Statistical mechanics of the anisotropic linear heisenberg model. 127(5):1508–1518.

[56] E. Barouch and B. M. McCoy. Statistical mechanics of the XY model. II. spin-correlation functions. 3(2):786–804.

[57] E. Barouch, B. M. McCoy, and M. Dresden. Statistical mechanics of the XY model. 2(3):1075–1092.

[58] E. Fradkin. *Field Theories Of Condensed Matter Systems Vol#82 On-demand Printing Of #52253*. CRC Press, 1 edition edition.

[59] N. N. Bogoljubov. On a new method in the theory of superconductivity. 7(6):794–805.

[60] F. A. Berezin. *The method of second quantization*. Number v. 24 in Pure and applied physics. Academic Press.

[61] B. Westwanski and A. Pawlikowski. On a general statistical wick theorem. 43(2):201–202.

[62] S. Cheong and C. L. Henley. Many-body density matrices for free fermions.

[63] C. V. Kraus, M. M. Wolf, J. I. Cirac, and G. Giedke. Pairing in fermionic systems: A quantum information perspective. 79(1):012306.

[64] M.A. Nielsen. The fermionic canonical commutation relations and the jordan-wigner transform. (6).

[65] A. Bashirov. Chapter 12 - fourier series and integrals. In Agamirza Bashirov, editor, *Mathematical Analysis Fundamentals*, pages 307–345. Elsevier.

[66] C. E. Shannon. A mathematical theory of communication. 27(3):379–423. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.1538-7305.1948.tb01338.x.

[67] D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 1 edition edition.

[68] R. G. Gallager. *Information Theory and Reliable Communication*.

[69] R. Gallager. The random coding bound is tight for the average code (corresp.). 19(2):244–246.

[70] C. P. Robert and G. Casella. The multi-stage gibbs sampler. In Christian P. Robert and George Casella, editors, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, pages 371–424. Springer.

[71] W. R. Gilks, S. Richardson, and D. Spiegelhalter. Markov chain monte carlo in practice.

[72] C. Seiler. Chapter 8 - bayesian statistics in computational anatomy. In Guoyan Zheng, Shuo Li, and Gabor Székely, editors, *Statistical Shape and Deformation Analysis*, pages 193 – 214. Academic Press, 2017.