Exploring Thermalization from Tipicallity's viewpoint in the XY model

An alternative mechanism for thermalization

Jose Alejandro Montaña Cortes

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

Exploring Thermalization from Tipicallity's viewpoint in the XY model

An alternative mechanism for thermalization

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

"*one can prove that infinitely many more initial states evolve after a long time towards a more uniform distribution of states than to a less uniform one, and that even in that latter case, these states will become uniform after an even longer time*" (L. Boltzmann 1866) [1]

Abstract

I think the abstract is going to be the last thing i am going to write
Key Words: Typicality, Thermalization, Ultra-Orthogonality..

Acknowledgments

Content

Chapter 1

Quantum Thermodynamics And Canonical Typicality

1.1 Introduction

The foundations of Statistical Mechanics remain a debatable subject. Fundamental questions have emerged from these discussions regarding the role of probabilities, entropy[1] and the relevance of time averages and ensemble averages to individual physical systems. One of the most controversial issue is the validity of the postulate of equal a priori probability, postulate which can not be proved.

In this chapter we are going to discuss some ideas based on typicality addresed by several authors [2–4], who have abandoned the unprovable aforementioned postulate and replace it with a new viewpoint, which is uniquely quantum, and which does not rely on any ignorance probabilities in the description of the state. Instead, is supported, and can be proved, by means of the entanglement between a physical system and it environment.

This chapter will be divided in three parts, first we are going to introduce the postulate of equal prior probability, discussing its commencement in the foundations of statistical mechanics, the idea of an ensemble as a collection of identical systems will be introduced and the postulate of equal prior probability will be translated to a particular version of ensemble, the microcanonical ensemble. Also a quantum version of the above-mentioned postulate will be exhibited in terms of the random phase postulate [5] and the derivation of the canonical ensemble for a weakly interacting system.

The second part will be dedicated to understand entanglement and therefore we present the phenomenon of canonical typicality, where thermalisation emerges as a consequence of typicality. These ideas are taken from *S.Popescu, A. J. short and A. Winter* [4, 6] work which we follow very close[2]. Quantitative arguments will be provided and explained with its

[1] As it is well known, probabilities and entropy are both referred as measures of ignorance.

[2] However, it is important to stress that an independet work done by *Goldstein et. al.* [3] discuss a similar issues addressed by *S.Popescu et. al.* in [4, 6].

respective mathematical tools needed to comprehend the ideas in which typicality is built on.
To conclude, we use all the previous results in the second part to present in the third part an approach studied by *Linden et. al.* [7] in which rather than a kinematic insight, they address the dynamical aspects of thermalisation, explaining a mechanism in which systems far from equilibrium would reach a generic state[3] [7]. Furthermore, we will talk about the main subject of this work, an idea which intuition is built on the ideas of typicality to show why in the case when a state can be decomposed as a linear superposition of states that have approximately the same energy, typicality will still holds and the random phase postulate could be replaced with something we call super-orthogonality over partial traces, and which can be understood in terms of an error correction scheme. therefore, it give us a mathematical proof of why thermalisation will still occurs in the case when we have states that have close eigenenergies.

1.2 Postulate of Equal a Priori Probability

Statistical physics have been historically inspired in phenomenological thermodynamics. In fact quantities such as temperature, pressure, heat and entropy emerged as concepts to describe empirical and macroscopical principles. the fact that we could describe a system by understanding how its microscopic constituents behave via some effective laws, known as *the law of thermodynamics [8]*[4]. A well known result from these studies was *the second law of thermodynamics* which let us know what kind of natural phenomena could occur.
It was not until Boltzmann, Gibbs and Maxwell that statistical physics became known as a fundamental theory, providing a bridge between experimental thermodynamics and the theoretical notions of the atomic world. However, this approach has been a subject of strong debates. Statistical physics differ from phenomenological thermodynamics by two main reasons. The first is the consideration that the behaviour of microscopic constituents could be described with classical dynamics, and the second is the introduction of probability and statistics to deduce the laws of thermodynamics.
We are going to introduce the main theoretical concepts proposed at the end of the 19th century, related to the description of the microscopic world[5].
Consider a mechanical system of N ($N \ggg 1$) particles subject to a time-independent po-

[3]Here a generic state has to be understood as a typical state, meaning the one that corresponds to its canonical state.

[4]For this very section we follow the ideas provided in [8] which turned up to be a very interesting historical review of statistical mechanics. This text is one of the most completed documented review in which not only the work of others is exposed but the controversy at the time related with these subjects. Even though is longer than i expected, it worth the time and i sincerely recommend to take a look at it.

[5]An interesting fact is that, it was Maxwell who really marks the rebirth of kinetic theory. His ideas about the how in a system with a large number of particles could be described taking a frequency interpretation of the probability was the base for Boltzmann's work.

tential V. Supposing every particle in the system has mass m and coordinates in its phase space $(\vec{q}_i, \vec{p}_i)$ we easily see that the state of the system is represented by a point in the phase space Γ $(\vec{q}, \vec{p}) = (\vec{q}_1, \ldots, \vec{q}_N, \vec{p}_1, \ldots, \vec{p}_N)$, and its time evolution follows the dynamics induced by its Hamiltonian. Moreover, imposing the restriction that the energy E is conserved, the evolution is confined to a set of Hamiltonians such that fulfill this restriction $(\Sigma_E = \{(\vec{q}, \vec{p}) \in \Gamma | H(\vec{q}, \vec{p}) = E\})$. At this moment is when probability comes to play an important role. In [1] Boltzmann proposed to interpret the probability associated to a particular state as the relative time spent by the system in that state[6]. This assumption relies on the hypothesis that the total time of a measurement is extremely long if compared with the intrinsic time scales of the evolution.
By Liouville's theorem, f is a constant function on all the admissible trajectories on Σ_E. Moreover, if we assume that the trajectory of a single point in phase space fills densely Σ_E (Ergodic hypothesis)

$$\rho(\vec{q}, \vec{p}) \propto \delta(H(\vec{q}, \vec{p}) - E). \tag{1.1}$$

based on this argument, Boltzmann was able to prove that thermal equilibrium could be described in terms of Maxwell's distribution of velocities.
In fact in both of his works [9], and [10], Boltzmann mentioned his dissatisfaction about the Ergodic hypothesis and slowly abandoned it. It seems that he only considered it an useful assumption for his general result

> For any system for which the hypothesis is true, its equilibrium state is characterized by[7],
>
> $$\rho_{\mathrm{mc}}\left(p|\vec{q}_1, \ldots, \vec{q}_N\right) dp = \frac{1}{\sqrt{2m\pi}} \frac{\Gamma\left(\frac{3N}{2}\right)}{\Gamma\left(\frac{3N-1}{2}\right)} \frac{\left(E - U - \frac{p^2}{2m}\right)^{(3N-2)/2}}{(E-U)^{(3N-3)/2}} dp, \tag{1.2}$$
>
> from which an analogy to the Maxwell distribution may be recovered in the limit $N \to \infty$, regardless of any details of the inter-particle interactions, or indeed whether the system represented is a gas, fluid, solid or any other thermal body.

1.2.1 Postulate of Equal a Priori Probability and the Microcanonical Ensemble

Al alternative path was taken by Gibbs who made use of an alternative idea of probability in what he named as "ensembles". In his book [11] Gibbs introduce the probability as a distribution function on a collection of identical systems instead of introducing the probability

[6] Concretely the probability of the phase point lies in an infinitesimal region of Σ_E is given by

$$\rho(q,p)\mathrm{d}q\mathrm{d}p = f(q,p)\delta(H(q,p) - E)\mathrm{d}q\mathrm{d}p,$$

where $dpdq$ is the Lebesgue measure of Γ and f is a suitable function.

[7] This equation was first derived by Boltzmann in [1], and this is the first occasion where probability considerations are applied to the state of the mechanical system as whole, instead of its individual particles.

as an ingredient associated to the state of a single system, as Boltzmann proposed. Gibbs considered "ensembles" in his pioneering treatment of statistical mechanics. His ensembles were infinite sets of macroscopically identical systems, each represented by a point in his phase space (also called *microstates*) being compatible with a single macroscopic configuration (*macrostate*). It is worth stressing that this approach is not interested in following the temporal evolution of a single microscopic configuration, but rather is concerned about the distribution of all the available microscopic configurations, even if the *microstate* associated, is entirely fictitious.
More precisely an ensemble is introduced as a probability density function on the phase space Γ, such that the average number of *microstates* in a region R of Γ is given by $\int_R \rho(\vec{q},\vec{p})d\vec{q}d\vec{p}$. Furthermore, the expectation value of an observable $f : \Gamma \to \mathbb{R}$ is the average of f over Γ

$$\langle f \rangle = \int_\Gamma f((\vec{q},\vec{p})\rho(\vec{q},\vec{p})d\vec{q}d\vec{p}. \tag{1.3}$$

In this approach we take the condition that the ensembles must be stationary[8] , to derive the microcanonical ensemble, which is the only compatible with the conservation of energy is such that

$$\rho_{\mu c}(\vec{q},\vec{p}) = \frac{1}{|\Sigma_E|}\delta(H(\vec{q},\vec{p}) - E), \tag{1.4}$$

with $|\Sigma_E|$ the measure of the set Σ_E.
As mentioned before Classical statistical mechanics is built on the postulate of Equal Probability. This result can be interpreted as follows:

> *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, when a macroscopic system is at equilibrium, every state compatible with the constrain of the system, is equally available (likely). Which formally is translated translates into the choice of a constant density function, called the microcanonical ensemble.

1.2.2 The Ergodic Hypothesis

We dedicate this special section to discuss a little more about the ergodic hypothesis and how Boltzmann could have understood this to give a formal argument of why it is needed that independent of the initial distribution of the system, this has to evolve towards a uniform distribution (equal a prior probability). First we will show some previous concepts to fully understand this ideas.
The ergodicity provides us an equality between ensemble and time averages of observables.

[8]This condition reads as

$$\frac{\partial \rho_t}{\partial t} = \{H, \rho_t\} = 0$$

Each measurement of an observable f at time t_0 takes a certain time to accomplish. In this period of time the observable f samples different values so that the effectively quantity is the time average

$$\frac{1}{t}\int_0^t f\left(T_s\left(\vec{q}_0,\vec{p}_0\right)\right)\mathrm{d}s, \tag{1.5}$$

with $(\vec{q}_0,\vec{p}_0)$ is the initial microstate (at $t=0$), and $\{T_s\}_{s\in\mathbb{R}}$ is the Hamiltonian flow generated by H[9]. Thus we are sampling f trajectories whose initial points are $(\vec{q}_0,\vec{p}_0)$. Furthermore, due to the difference in scales of time between the measurement and the intrinsic time of the microscopical constituents, it is possible to consider the limit

$$f^*\left(\vec{q}_0,\vec{p}_0\right)=\lim_{t\to\infty}\frac{1}{t}\int_0^t f\left(T_s\left(q_0,p_0\right)\right)\mathrm{d}s. \tag{1.6}$$

The problem about the ergodic hypothesis is to determine when the average ensemble average equals the time averages. Of course the general answer to this question is that this equality does not always holds, since $f^*(\vec{q}_0,\vec{p}_0)$ depends on the initial conditions whereas $\langle f\rangle$ does not. nevertheless these differences, the cases when the equality between time and ensemble averages holds, we say that the ergodic hypothesis is valid.

Perhaps the most well-known, interpretation to this problem is the one provided by the Ehrenfests [12]. In essence, they suggest that Boltzmann somehow relied on the ergodic hypothesis in his argument, even though he never explicitly mentioned it.

It is indeed evident that if the ergodic hypothesis holds, a state will spend time in the various regions of the energy hypersurface in phase space in proportion to their volume. In other words, during the evolution of the system along its trajectory, regions with small volume, corresponding to highly non-uniform distributions of state, are visited only sporadically, and regions with larger volume, corresponding to more uniform distributions of state, more often. Therefore, the latter could also make plausible the idea that if a system starts out from a very small region (a very improbable state), it will display a tendency to evolve towards the overwhelmingly larger equilibrium state.

Thus, as is mention in [12], it is suggested that Boltzmann relied on the ergodic hypothesis in order to equate time averages and phase averages, or in other words, to equate two meanings of probability (relative time and relative volume in phase space.) There is however no evidence that Boltzmann ever followed this line of reasoning neither in the 1870s, nor later. He simply never gave any justification for equivocating time and particle averages, or phase averages, at all. Presumably, he thought nothing much depended on this issue and that it was a matter of taste [10].

[9]Here H is the Hamiltonian function describing the system of N particles subject to a time-independet potential V

$$H(\vec{q},\vec{p})=\frac{1}{2m}\sum_{i=1}^{N}\vec{p^2}_i+V\left(\vec{q}_1,\ldots,\vec{q}_N\right).$$

[10]For further readings on ergodic hypothesis i sincerely recommend [8, 13–16]

1.2.3 The Quantum Postulate of Equal a Priori Probability

The latter discussion is completely classical. Since our world is quantum, a quantum formulation of statistical mechanics is needed. Instead of probability distributions that live in phase space, one should consider density matrices, which encode the whole physical content of the system. Therefore, the way we write the postulate of equal a priori probability has to be adapted to the formalism of quantum mechanics.
In quantum mechanics a system is described in a Hilbert space $\mathcal{H}$ and its evolution is generated by a Hamiltonian operator $\hat{H}$. By fixing the energy to belong to an energy shell around the value E ($[E, E+\delta]$), with $\delta \ll E$, but δ large enough so that the shell contains many energy eigenvalues of $\hat{H}$. Rather than considering a phase space region Γ_E, as we did in the classical case, we consider a subspace spanned by all eigenvectors with energy eigenvalues belonging to the energy shell ($\mathcal{H}_R = \mathcal{H}_{[E,E+\delta]}$).
The postulate of equal prior probability affirms that all the states compatible with the energy E are equiprobable[11]. However, in the quantum case this is not enough. In fact, these states must be in an incoherent superposition. This what is know as *the random phase postulate* and its a contribution to the foundations of statistical mechanics which is purely quantum.

The idea here is to assume that the coupling between the system and its environment is sufficiently weak such that the energy of the system is found with in a macroscopically narrow range $[E, E+\delta]$ containing many possible states of the system. Transitions between this energy are mediated by ΔH. All states within energy range which can be connected by range which can be connected by ΔH are considered accessible. We assume that the environment is sufficiently complex, its states so numerous, and its transitions so rapid that phase relationships between different states of the system can not be maintained over microscopically long time intervals. Formally the idea is then that the Fourier coefficient of a vector state in $\mathcal{H}_R$ should have equal probability and completely random phases, due to the unavoidable interactions between the environment and the system [17].
Using these two postulates we conclude that the state of the universe is described in terms of a projector of the Hamiltonian $\hat{H}$ in the constrained Hilbert space $\mathcal{H}_R$. We define $\mathcal{E}_R$, the equiprobable state of the universe corresponding to the restriction R by

$$\mathcal{E}_R = \frac{\mathbb{I}_R}{d_R}, \tag{1.7}$$

where $\mathbb{I}_R$ is the identity (projection) operator on $\mathcal{H}_R$, and d_R is the dimension of $\mathcal{H}_R$. $\mathcal{E}_R$ is the maximally mixed state in $\mathcal{H}_R$, in which each pure state has equal probability. This corresponds to the standard intuition of assigning equal a priori probabilities to all states of the universe consistent with the constraints.

[11] In the quantum mechanical frame this reads as every eigenstate belonging to $\mathcal{H}_R$

1.2.4 The canonical Ensemble

In this part we will show that under some quite general assumptions it is possible to show that if the universe is in the equiprobable state, every small part of it (system) will be in a canonical state Ω_S characterised by a Boltzmann distribution among its eigenstates at a given temperate β^{-1}, this is

$$\Omega_S \propto e^{-\beta H_S} \tag{1.8}$$

To prove this we make a division of the universe in two parts. The system S and its environment E. So Hilbert space will then be described by $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$. Then the Hamiltonian of the universe will be written as

$$\hat{H} = \hat{H}_S \otimes \hat{\mathbb{I}}_E + \hat{\mathbb{I}}_S \otimes \hat{H}_E + \hat{H}_{\text{int}}, \tag{1.9}$$

where $\hat{H}_S$ and $\hat{H}_E$ act separately on the system and its environment respectively, $\hat{\mathbb{I}}_S$ and $\hat{\mathbb{I}}_E$ are the identity operators on $\mathcal{H}_S$ and $\mathcal{H}_E$, and $hat H_{\text{int}}$ describes the interaction between the system and its environment. Supposing the interaction is weak[12] and assuming the dimension of $\mathcal{H}_E$ is much larger than the dimension of $\mathcal{H}_S$ ($d_E = \dim \mathcal{H}_E \gg d_S = \dim \mathcal{H}_S$). We suppose the macroscopic energy E belongs to a small energy interval ($[E, E+\delta]$)[13]. Let $\mathbb{I}_R$ be the identity (projector) operator of $\hat{H}$ on the energy shell

$$\mathcal{H}_R = \mathcal{H}_{[E,E+\delta]}. \tag{1.10}$$

Assuming the universe is in the equiprobable state $\mathcal{E}_R$ given by (1.7). The state of the system S can be obtained by partial trace the state of the universe over the environment, that is,

$$\Omega_S = \text{Tr}_E \, \mathcal{R}. \tag{1.11}$$

From now on the main idea will be to show that Ω_S corresponds to the thermal state at a given temperature[14].
Let $\{|E_k\rangle\}_{k=1}^{d_E} \subset \mathcal{H}_E$ and $\{|\varepsilon_\alpha\rangle\}_{k=1}^{d_S} \subset \mathcal{H}_S$ be the energy basis of $\hat{H}_E$ and $\hat{H}_S$ respectively[15]. Since we use the hypothesis that the interaction is weak, we have that the equiprobable state

[12] This can be understood in terms of the Hamiltonians as

$$||\hat{H}_{\text{int}}|| \lll ||\hat{H}_S||, ||H_E||.$$

[13] As mentioned before here we suppose $\delta \lll E$ on a macroscopic scale but large enough to contain many eigenvalues on H_E

[14] The following prove was taken from [17]

[15] In other words this mean that the Hilbert $\mathcal{H}$ space associated with the universe has basis given by

$$\{|\varepsilon_\alpha\rangle \otimes |E_k\rangle : 1 \leq \alpha \leq d_S, 1 \leq k \leq d_E\}$$

$\mathcal{E}_R$ is written as[16].

$$\mathcal{E}_R = \frac{\mathbb{I}_R}{d_R} \approx \frac{1}{d_R} \sum_{\alpha,k} \chi_{[E,E+\delta]} (\varepsilon_\alpha + E_k) |\varepsilon_\alpha\rangle \langle\varepsilon_\alpha| \otimes |E_k\rangle \langle E_k| . \tag{1.12}$$

By tracing over the environment we get

$$\Omega_S = \mathrm{Tr}_E \, \mathcal{E}_R = \frac{1}{d_R} \sum_{\alpha,k} \chi_{[E,E+\delta]} (\varepsilon_\alpha + E_k) |\varepsilon_\alpha\rangle \langle\varepsilon_\alpha| = \frac{1}{d_R} \sum_{\alpha} d_\alpha^{(E)} |\varepsilon_\alpha\rangle \langle\varepsilon_\alpha| , \tag{1.13}$$

where

$$d_\alpha^{(E)} = \sum_k \chi_{[E,E+\delta]} (\varepsilon_\alpha + E_k) = \sum_k \chi_{[E-\varepsilon_\alpha,E-\varepsilon_\alpha+\delta]} (E_k) . \tag{1.14}$$

Since $\hat{H}_E = \sum_k E_k |E_k\rangle \langle E_k|$, we get that

$$d_\alpha^{(E)} = \mathrm{Tr} \, \chi_{[E-\varepsilon_\alpha,E-\varepsilon_\alpha+\delta]} (H_B) = \dim \mathcal{H}^{(E)}_{[E-\varepsilon_\alpha,E-\varepsilon_\alpha+\delta]}, \tag{1.15}$$

where $\mathcal{H}^{(E)}_{[E_1,E_2]} \subset \mathcal{H}_E$ is the subspace generated by all eigenstates with energy in $[E_1, E_2]$. Thus $d_\alpha^{(E)}$ is a non-negative integer.
Defining the bath entropy at energy E as the logarithm of the number on energy levels in the bath,

$$S_E(E) = \ln \left(\dim \mathcal{H}^{(E)}_{[E,E+\delta]} \right) , \tag{1.16}$$

we deduce,

$$d_\alpha^{(E)} = \dim \mathcal{H}^{(E)}_{[E-\varepsilon_\alpha,E-\varepsilon_\alpha+\delta]} = \mathrm{e}^{S_E(E-\varepsilon_\alpha)}. \tag{1.17}$$

As the dimension of $\mathcal{H}_E$ is very large, $d_R \ggg 1$, we can therefore assume that the spectrum of $\hat{H}_E$ is quasi-continuous, so that $S_E(E)$ can be considered a continuous differentiable function of E. If we assume that the microscopic energy is much smaller than the macroscopic energy E ($\varepsilon \lll E$), we can write

$$S_E (E - \varepsilon_\alpha) \approx S_E(E) - \frac{\mathrm{d}S_E(E)}{\mathrm{d}E} \varepsilon_\alpha \tag{1.18}$$

Thus, we get

$$\Omega_S = \frac{1}{d_R} \sum_{\alpha} d_\alpha^{(E)} |\varepsilon_\alpha\rangle \langle\varepsilon_\alpha| \approx \frac{1}{Z} \sum_{\alpha} \mathrm{e}^{-\beta\varepsilon_\alpha} |\varepsilon_\alpha\rangle \langle\varepsilon_\alpha| = \frac{1}{Z} \mathrm{e}^{-\beta \hat{H}_S}, \tag{1.19}$$

with $Z = \mathrm{Tr} \left(e^{-\beta \hat{H}_S} \right)$, and

$$\beta = \frac{\mathrm{d}S_E(E)}{\mathrm{d}E}. \tag{1.20}$$

is the thermodynamical expression of the inverse temperature of the bath.

[16]Here the sums is over indexes k and α such that $\varepsilon_\alpha + E_k \in [E, E + \delta]$. Which is equivalent to say where the characteristic function does not vanish

1.3 Entanglement and Canonical Typicality

As mentioned before we dedicate this section to show how the postulate of equal a priori probability can be proved rather than postulated as we quoted before. The most interesting thing is that this postulate emerges from the very structure of quantum mechanics and it is a simply consequence of entanglement.

1.3.1 Entanglement and the Foundations of Statistical Mechanics

> When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz. by endowing each of them with a representative of its own (E. Schrödinger 1995 [18]).

To understand what Schrödinger meant by this, we can illustrate the situation by considering a composite system of two spins on the Hilbert space $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$, where $\mathcal{H}_S = \mathcal{H}_E = \mathbb{C}^2$. Denoting $\{|\uparrow\rangle, |\downarrow\rangle\}$, our computational basis of $\mathbb{C}^2$ [19–21]. If the bipartite system is described in terms of a factorized state, for example,

$$|\phi\rangle = |\uparrow\rangle_S \otimes |\uparrow\rangle_E, \tag{1.21}$$

then the system and its environment can be described independently. On the contrary, when the global state is not factorized (Bell state),

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle_S \otimes |\uparrow\rangle_E + |\downarrow\rangle_S \otimes |\downarrow\rangle_E\right) \tag{1.22}$$

then the state is entangled. In fact the density matrix associated with this system is

$$\rho_S = \mathrm{Tr}_E \left(\left|\Phi^+\right\rangle \left\langle\Phi^+\right|\right) = \frac{1}{2}\hat{\mathbb{I}}_2 \tag{1.23}$$

where $\hat{\mathbb{I}}$ is the identity operator on $\mathcal{H}_S = \mathbb{C}^2$. the latter equation tells us that the system S is in a totally mixed state, that is, it is uniformly random distributed in $\mathcal{H}$.
Now consider the general parametrization of the latter example,

$$|\Phi_\alpha\rangle = \sqrt{\alpha}|\uparrow\rangle_S \otimes |\uparrow\rangle_E + \sqrt{1-\alpha}|\downarrow\rangle_S \otimes |\downarrow\rangle_E, \quad \alpha \in [0,1], \tag{1.24}$$

which condensate the separable state $|\phi\rangle$ ($\alpha = 0$), and the Bell state$|\Phi^+\rangle = \left|\Phi_{1/2}\right\rangle$ ($\alpha = 1/2$). The reduced density matrix is then given by

$$\rho_S^\alpha = \alpha|\uparrow\rangle\langle\uparrow| + (1-\alpha)|\downarrow\rangle\langle\downarrow|. \tag{1.25}$$

which is a state whose mixture depends on α.
Summarising, if the state of the composite system is factorised, the information on the whole

state and on every subsystem is completely accessible. On the other hand, if the global state is entangled, even though we have complete knowledge of the state in the universe, a prior only a partial knowledge of the subsystem can be obtained. Particularly, if the global state is maximally entangled ($\alpha = 1/2$) one has no information at all on the subsystem. Mathematically speaking, the information content of a state ρ is described by the von Neumann entropy [22].

$$S(\rho) = -\operatorname{tr}(\rho \ln \rho) = -\sum_k p_k \ln p_k, \tag{1.26}$$

where p_k are the eigenvalues of ρ, subject to the conditions

- $0 \leq p_k \leq 1$,
- $\sum_k p_k = 1$.

It is straightforward to realise that every pure state has 0 entropy, which means that the information encoded in the state is completely available. In our case, for example we see that $S(|\Phi_\alpha\rangle \langle\Phi_\alpha|) = 0$ for every α. And contrarily, for the general parametrised reduced state we get

$$S\left(\rho_S^\alpha\right) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha), \tag{1.27}$$

that is a positive metric function of $\alpha \in [0, 1]$, which is 0 for separable global states ($\alpha = 0, 1$)

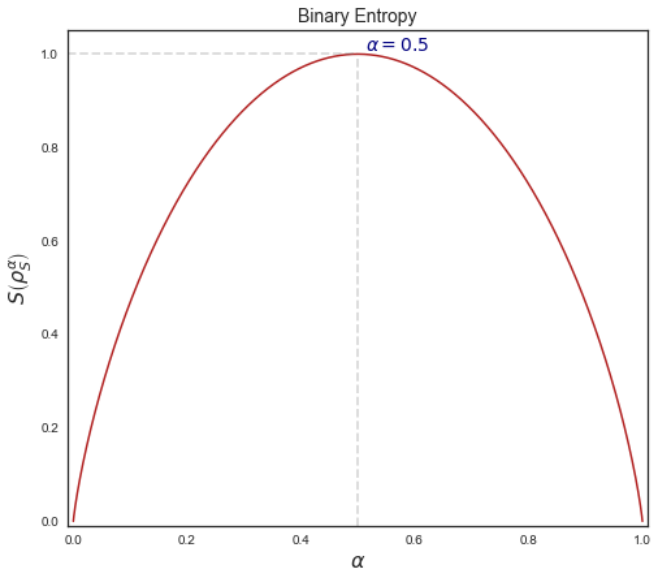


Figure 1.1: Binary Entropy as a function of binary outcome probability α.

and reaches its maximum for the maximally entangled Bell state ($\alpha = 1/2$). In this latter example the entropy is maximal and it corresponds to a complete ignorance on the subsystem

S. Observe that the expression given in (1.27) is nothing but the binary entropy function [23] which is equivalent to the Shannon entropy [24] of the probability vector $(\alpha, 1-\alpha)$.
Form the latter example it is easy to generalise it to a given pure state $|\Psi\rangle$ of a composite system $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_E$ with generic dimensions $d_S = \dim \mathcal{H}_S \leq d_E = \dim \mathcal{H}_E$, one gets that

$$0 \leq S(\rho_S) \leq \ln d_S. \tag{1.28}$$

Here $S(\rho_S) = 0$ for separable state, $|\Psi\rangle = |u\rangle \otimes |v\rangle$, while $S(\rho_S) = \ln d_S$ for maximally entangled states,

$$|\Psi\rangle = \frac{1}{\sqrt{d_S}} \sum_{k=1}^{d_S} |u_k\rangle \otimes |v_k\rangle \,, \tag{1.29}$$

with $\{u_k\}, \{v_k\}$ are the orthogonal basis for each of system and environment. The von Neumann entropy is a measure of entanglement which leads to an objective lack of knowledge[17]. In fact, this objective "lack of knowledge" is related to the state of the system, because even if we knew everything about the state of the universe, every subsystem (small portion of the universe) could be mixed [26,27]. This is something astonishing since this lack of knowledge came form nothing but the true nature of quantum mechanics, and no randomness is introduced[18].
In the following part we will show that thermalisation appear as a generic property of pure states of the universe, meaning that for the overwhelming majority of them, the reduced state of the system is the canonical mixed state. As a conclusion we can state that the postulate of equal a prior probability, which refers to ensembles or time averages of states of the universe, and as such relies on a subjective lack of information, can be disregard and one can refer only to pure state of the universe. The lack of information which will give a canonical density matrix for the system is just a physical consequence of entanglement between the system and its environment.

1.3.2 Canonical Typicality

In this part we are going to show that the principle of equal prior probability, which can not be proved, should be replaced with the principle of *Canonical typicality*, which is based on individual states rather than ensembles or time averages, and most importantly, can be

[17] In general entropy provides a tool that can be used to quantify entanglement, although other entanglement measures exist [25] If the overall system is pure, the entropy of one subsystem can be used to measure its degree of entanglement with the other subsystems. In particular for bipartite pure states, the von Neumann entropy of reduced states is the unique measure of entanglement in the sense that it is the only function on the family of states that satisfies certain axioms required of an entanglement measure. Being this the reason why we only introduce this entropy.

[18] Classically the complete knowledge of the state of the universe implies a complete knowledge of the state of any subsystem. Understanding why in the case of classical mechanics, randomness was artificially add to the description.

proved.
Consider a large quantum mechanical system, "The Universe", which we decompose in two parts, the system S and the environment E[19]. 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

$$\mathcal{H}_R \subset \mathcal{H}_S \otimes \mathcal{H}_E \tag{1.30}$$

As we mentioned before, the dimensions of $\mathcal{H}_S$, $\mathcal{H}_E$ and $\mathcal{H}_R$ will be denote by d_S, d_E and d_R respectively. In the standard approach to statistical mechanics, as we aforementioned, the restriction is imposed on the total energy. However, as *Popescu et. al.* mention in [4, 6], this restriction can be let to be completely arbitrary.
Using the definition of the equiprobable state in (1.7), we know that $\mathcal{E}_R$ is the maximally mixed state in $\mathcal{H}_R$, in which pure state has equal probability[20].
The *canonical state* of the system S is defined as the trace over the environment,

$$\Omega_S = \mathrm{Tr}_E\, \mathcal{E}_R. \tag{1.31}$$

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

$$\rho_S = \mathrm{tr}_B(|\phi\rangle\langle\phi|) \tag{1.32}$$

Here we ask ourselves, how different is ρ_S form the canonical state Ω_S. The answer to this is provided by *Popescu et. al.* in [4, 6], which states that ρ_S is very close to Ω_S for 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 $\mathcal{E}_R$.
Nevertheless, it is important to stress that Ω_S is not necessarily the thermal canonical state (1.19), but rather a generalised canonical state with respect to the arbitrary restriction R chosen[21].
As an illustrative example, we can interpret this result by representing our universe obeying the global constraint R as a map chart in which we see nothing but a vast ocean representing the equiprobable states in $\mathcal{H}_R$ and some islands representing those states who differ drastically from the canonical state. Particularly we represent with a boat a random picked state in the universe, as one can see, for the vast majority of random choices we have, this boat would land in a portion of this ocean.

[19] Here it is implicit that the dimension of the environment is much larger than the one of the system.

[20] As Popescu mention in his paper, this corresponds to the standard intuition in classical mechanics of assigning equal a priori probabilities to all states of the universe consistent with the constraints [4, 6].

[21] The thermodynamic interpretation is recover is the restriction imposed coincide with the total energy. In that case we can state that almost every pure state $|\phi\rangle$ of the universe is such that the system S is approximately in the canonical thermal state $e^{-\beta \hat{H}_S}/Z$.

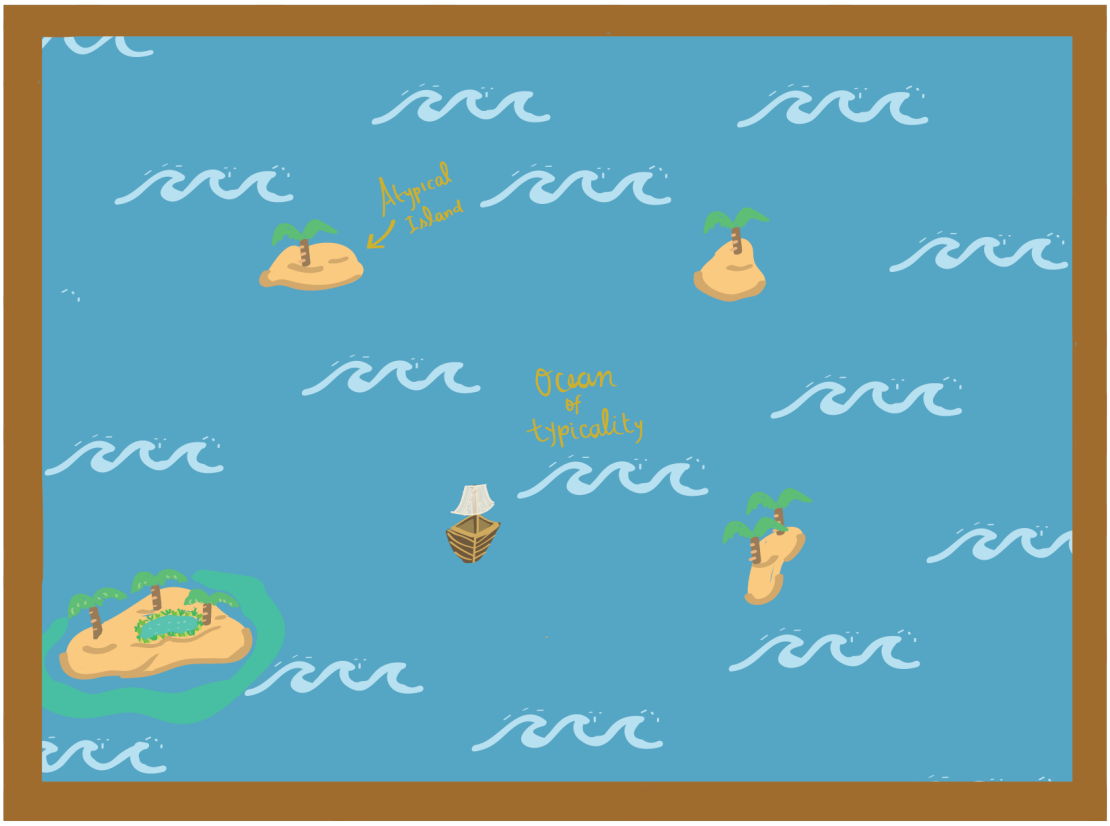


Figure 1.2: Illustration of the typicality in a map chart. The boat here refers to a random picked state in our universe. Here the islands refer to those states who radically differ from the Canonical state.

1.3.3 Quantitative Arguments behind Typicality

In order to formally express these ideas, it is important provide more details. To start, it is necessary to first define a notion of distance between states ρ_S and Ω_S, as well as a measure over which pure states $|\phi\rangle$ are defined.
Defining the distance of two operators is not a trivial task. As we will see depending on the distance we choose connections to a physical quantity could emerge. We start by defining the trace distance between ρ_S and the canonical state Ω_S , by $||\rho_S - \Omega||_1$, which is defined via the trace norm

$$||\rho||_1 = \mathrm{Tr}\,|\rho| = \mathrm{Tr}\left(\sqrt{\rho^\dagger \rho}\right) \tag{1.33}$$

An interesting fact about this distance is that it represents the maximal difference in probability of obtaining any outcome for any measurement performed on the states ρ_S and Ω_S. This can be clearly seen if we write explicitly the definition of an induced norm

$$\|\rho\| = \sup_{\|M\|=1} |\,\mathrm{Tr}(\rho M)|. \tag{1.34}$$

Using this, we can say that the trace distance measures how difficult is to know ρ_S and Ω_S via a measurement M, explicitly this is

$$|\mathrm{Tr}\,(\rho_S M) - \mathrm{Tr}\,(\Omega_S M)| \leq \|\rho_S - \Omega_S\|_1 \, \|M\|, \tag{1.35}$$

An important fact to stress is the choice of distance, one could wonder, why do not we use other operator distance such like Hilbert-Schmidt norm, which is defined by taking the root square of the trace, instead of the trace of the root square, as the trace norm is defined. The reason to do so is simply because the Hilbert-Schmidt distance in higher dimensions can tend to be very small even when the states have a disjoint support [28]. To clarify this, consider in $\mathbb{C}^{2d}$ the two states $\rho_1 = \mathbb{I}_1/d$ and $\rho_2 = (1 - \mathbb{I}_1)/d$, where $\mathbb{I}_1$ is a projector over d dimensions. An important fact over these states is that they have disjoint supports just as we mentioned before. So computing the distance between these two states give us

$$\|\rho_1 - \rho_2\|_1 = 2, \quad \|\rho_1 - \rho_2\|_2 = \sqrt{\frac{2}{d}}, \tag{1.36}$$

and then the trace norm give us a constant while the Hilbert-Schmidt distance decreases d increases $1/\sqrt{d}$.
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 the real sphere $(2d_R-1)$-dimensional. Therefore, the states we are interested on, live over the surface of a sphere of d_R dimensions, thus by random sample one of this states we will have to sample with the measure $\sigma(\mathbb{S}^{2d_R-1})$, which is known as the Haar measure. From this definition, it is clear to see if the selected random state over the Haar measure, the average state of the universe in $\mathcal{H}_R$ is nothing but the equiprobable state, and then we get that average state of the system is nothing but the canonical state $\Omega_S = \langle\rho_S\rangle$.

Now with the notion of the distance we chose to work with, and the space where these pure states live in, we are ready to announce the general result in typicality.

Theorem of Canonical Typicality [4, 6].

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:

$$\mathrm{Prob}\,(\|\rho_S - \Omega_S\|_1 \geq \eta) \leq \eta', \tag{1.37}$$

where

$$\eta = \varepsilon + \sqrt{\frac{d_S}{d_E^{\mathrm{eff}}}}, \quad \eta' = 2\exp\left(-C d_R \varepsilon^2\right), \tag{1.38}$$

with

$$C = \frac{1}{18\pi^3}, \quad d_E^{\text{eff}} = \frac{1}{\operatorname{Tr}\Omega_E^2} \geq \frac{d_R}{d_S}, \quad \Omega_E = \operatorname{Tr}_S \mathcal{E}_R \tag{1.39}$$

Form the latter result we notice that η and η' are small quantities, and then, the state will be close to the canonical state with high probability, whenever $d_E^{\text{eff}} \ggg d_S$ and $d_R\varepsilon^2 \ggg 1 \ggg \varepsilon$. The latter condition can be ensured when $d_R \ggg 1$.

So what our latter results holds 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, we will have always (almost) that every system, with small enough dimension, will be indistinguishable from the canonical state. Even more, *Popescu et. al.* get an expression to show how the fluctuations around the average behave. For this they get the next result

$$\langle \|\rho_S - \Omega_S\|_1 \rangle \leq \sqrt{\frac{d_S}{d_E^{\text{eff}}}} \leq \sqrt{\frac{d_S^2}{d_R}}, \tag{1.40}$$

which tells us, if the dimension of our accessible space d_R is large enough, compared to the dimension of the system, the fluctuations on the system are also very small. Thus these two results give us the quantitative arguments of typicality.

Up to this moment we have provided the qualitative as well quantitative arguments to show why typicality gives us an approach different than the equal a priori probability, that can not be proved, and replace it with a key quantum property, entanglement, which provides us a subjective lack of knowledge and more important that can be proved, to show that thermalisation turns up to be a generic property of pure states of the universe. However, despite this result explains very well the reason why by choosing randomly 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 (atypical state) under certain unitary evolution can reach its thermalisation. This means that the latter result is Kinematic rather than dynamical and therefore can not be used to explain how a thermalisation in a system can occur. The goal of the next section is to introduce a work done by *Linden et. al.* [7] in which a dynamical approach is done in order to explain how thermalisation can occur in a system reliant on a given unitary dynamic.

1.4 Evolution Towards Equilibrium.

As we above-mentioned, the latter result is only valid for a given time and state, meaning that this can not be used to study the how thermalisation occurs. Now we are interested in states that are atypical, in the sense that are going to be those states such that drastically differ from the canonical state. Even though, we might think that if typicality holds, most

evolutions will quickly drive us from a state in which the system is not thermalised into one that is, and that system will remain in this states of "thermalisation" for most of its evolution. Nevertheless, there are some problems which are worth to mention about why this is much harder problem to solve. In the following we will show what could be named as the roadmap of what is needed to be proved in order to show that a thermal state has thermalised.

- **Equilibration**: It is possible to affirm that a system will equilibrate if its states evolve towards a particular state, which can be in the more general case a mixed state, and remains in that state, or at least quite close to it, for every time[22]
- **Environment state independence**: The equilibrium state that the system reaches has to be independent of the initial state of the environment, this is, when the system reaches its state of equilibrium, this state should depend only on macroscopic parameters of the environment, like temperature or similar macroscopic parameters.
- **System state independence**: If the system is much smaller than the environment, the state of equilibrium should not depend of its initial state.

Having this in mind, it is possible to tackle these problems one by one. An interesting result is provided by *Linden et. al.* [7], where they manage to prove that with relatively full generality, equilibrium is an universal property of Quantum systems and even more that equilibrium state does not depend on the state of the environment.
To completely understand *Linden et. al.* work is necessary to first understand couple of definitions they provide in their paper [7].
Universe: Here we will refer always to a large quantum universe living in a Hilbert space $\mathcal{H}$. As previously, we are considering an universe that can be decomposed in two, in this decomposition we refer the system S to a small part of the Hilbert space and the rest we will call it the environment. Explicitly we 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 the respective dimension of the system and the environment. Notice that here the environment nor the system have provided with any special property. Here the system could be a single particle or even a section of a lattice.
Hamiltonian: The evolution of the universe will be governed by a Hamiltonian given by

$$\hat{H} = \sum_k E_k \left|E_k\right\rangle \left\langle E_k\right| . \tag{1.41}$$

with $|E_k\rangle$ the eigenstate in the energy basis with energy E_k. The main assumption here is related with the possible values of energies this Hamiltonian can have. The only requirement

[22]Even though this definition does not specify the sort of equilibrium state of the system, the state of equilibrium will strongly depend on the initial condition of the system as well as the initial conditions over its environment.

needed the Hamiltonian to have non-degenerate energy gaps.
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 $\ell = n$. Which turns out to be the same condition of imposing that the energy levels have to be non-degenerate.
Notice that the restriction imposed to the Hamiltonian is a extremely natural constraint, this is because all Hamiltonian that lack of symmetries have non-degenerate energies, so we refer to a set of Hamiltonians with measure 1.
Notation: We will work here with pure states for the universe represented by $|\Psi(t)\rangle$ with a density matrix state given by $\rho(r) = |\Psi(t)\rangle \langle\Psi(t)|$. As above-mentioned, the state of the system is obtained by tracing out the environment at a time t, that is, $\rho_S(t) = \mathrm{Tr}_E\, \rho(t)$. Similarly we define the state of the environment as $\rho_E(t) = \mathrm{Tr}_S\, \rho(t)$.
We define a convenient quantity which is the transient state, or the time averaged state ω

$$\omega = \langle \rho(t) \rangle_t = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau \rho(t) dt, \tag{1.42}$$

and similarly we define ω_s and ω_E as the time averaged state of the system and the environment respectively.
It is also convenient to re introduce a concept we have already used, the effective dimension of a mixed state ρ:

$$d^{\text{eff}}(\rho) = \frac{1}{\mathrm{Tr}\,(\rho^2)}, \tag{1.43}$$

which is generally a better measurement of the effective dimension than the dimension of the support of ρ. This measure roughly tells us how many states contribute to the mixture[23], carrying the probabilistic weight of different states in the mixture, and is a continuous measure.
With the concepts aforementioned Linden et. al. [7] are able to mathematically prove

> *Every pure state of a quantum universe, composed by a large number of eigenstates of energy*[24] *such that evolves under an arbitrary Hamiltonian, is such that every system small enough will equilibrate.*

The reason why the latter statement requires the universe to have many changes in its time evolution, is because for equilibration to take place it is needed that part of the information of the initial state of the system leaves the system and enters in the environment. This notion of evolving through many states can be mathematically encapsulated via the effective

[23] Particularly, a mixture of n orthogonal states with equal probability has effective dimension of n

[24] The reason why we need the global state to have many eigenstates of energy is because by imposing this we can assure that there will be a large quantity of changes throughout the evolution of the system.

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 $|Psi(t)\rangle$ as

$$|\Psi(t)\rangle = \sum_k c_k e^{-iE_k t} |E_k\rangle \tag{1.44}$$

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

$$\rho(t) = \sum_{k,l} c_k c_l^* e^{-i(E_k - E_l)t} |E_k\rangle \langle E_l| , \tag{1.45}$$

which can be expanded and written as

$$\begin{aligned} \rho(t) &= \underbrace{\sum_n \|c_n\|^2 |E_n\rangle \langle E_n|}_{\omega} + \underbrace{\sum_{m\neq n} c_n c_m^* |E_n\rangle \langle E_m| e^{-it(E_n - E_m)}}_{\lambda(t)} \\ &= \omega + \lambda(t), \end{aligned} \tag{1.46}$$

Now in the case of non-degeneracy of the energy levels we have

$$\omega = \langle\rho(t)\rangle_t = \sum_k |c_k|^2 |E_k\rangle \langle E_k| , \tag{1.47}$$

leading us to

$$d^{\text{eff}}(\omega) = \frac{1}{\text{Tr}\,(\omega^2)} = \frac{1}{\sum_k |c_k|^4}. \tag{1.48}$$

Thus formally the statement above-mentioned can be mathematically written in terms of central quantity $D\left[\rho_S(t), \omega_S\right]$[25], the distance between $\rho_S(t)$, the state of the system at a time t, and its time average, $\omega_S = \langle\rho_S(t)\rangle_t$. The difference between $\rho_S(t)$ and ω_S in the energy eigenstates can be written as

$$\rho_S(t) - \omega_S = \sum_{m\neq n} c_m c_n^* e^{-i(E_m - E_n)t} \,\text{Tr}_E\, |E_m\rangle \langle E_n| . \tag{1.49}$$

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. Therefore, in order to characterise these fluctuations we will be interested in the time average of distance $\langle D\left[\rho_S(t), \omega_S\right]\rangle_t$, so the value this average takes will tell us about where the system is spending most of its time. In other words $\langle D\left[\rho_S(t), \omega_S\right]\rangle_t$ will be small when the system equilibrates to ω_S.
To be able to prove what is announced as the *Theorem 1* in [7] it is useful to relate the trace distance to the square of the Hilbert-Schmidt distance using a standard bound provided in [29]

$$D\left(\rho_1, \rho_2\right) = \frac{1}{2} \,\text{Tr}_S \sqrt{\left(\rho_1 - \rho_2\right)^2} \leq \frac{1}{2} \sqrt{d_S \,\text{Tr}_S \left(\rho_1 - \rho_2\right)^2}. \tag{1.50}$$

[25]Here has in the case of typicality we use the trace distance.

By using the concavity of the square-root function, we therefore have

$$\langle D\left[\rho_S(t),\omega_S\right]\rangle_t \leq \sqrt{d_S\left\langle \mathrm{Tr}_S\left[\rho_S(t)-\omega_S\right]^2\right\rangle_t}, \tag{1.51}$$

which will provide us the bound we need to proof the theorem. Now using (1.49) we write

$$\left\langle \mathrm{Tr}_{\mathcal{S}}\left[\rho_{\mathcal{S}}(t)-\omega_{\mathcal{S}}\right]^2\right\rangle_t = \sum_{m\neq n}\sum_{k\neq l}\mathcal{T}_{klmn}\,\mathrm{Tr}_{\mathcal{S}}\left(\mathrm{Tr}_E\left|E_k\right\rangle\left\langle E_l\right|\mathrm{Tr}_E\left|E_m\right\rangle\left\langle E_n\right|\right), \tag{1.52}$$

where $\mathcal{T}_{klmn} = c_k c_l^* c_m c_n^* e^{-i(E_k-E_l+E_m-E_n)t}$. Computing this time average taking into account that the Hamiltonian has non-degenerate energy gaps[26] we find that

$$\begin{aligned}
\left\langle \mathrm{Tr}_S\left[\rho_S(t)-\omega_S\right]^2\right\rangle_t &= \sum_{k\neq l}\left|c_k\right|^2\left|c_l\right|^2\mathrm{Tr}_S\left(\mathrm{Tr}_E\left|E_k\right\rangle\left\langle E_l\right|\mathrm{Tr}_E\left|E_l\right\rangle\left\langle E_k\right|\right)\\
&= \sum_{k\neq l}\left|c_k\right|^2\left|c_l\right|^2\sum_{ss'bb'}\left\langle sb|E_k\right\rangle\left\langle E_l|s'b\right\rangle\left\langle s'b'|E_l\right\rangle\left\langle E_k|sb'\right\rangle\\
&= \sum_{k\neq l}\left|c_k\right|^2\left|c_l\right|^2\sum_{ss'bb'}\left\langle sb|E_k\right\rangle\left\langle E_k|sb'\right\rangle\left\langle s'b'|E_l\right\rangle\left\langle E_l|s'b\right\rangle\\
&= \sum_{k\neq l}\left|c_k\right|^2\left|c_l\right|^2\mathrm{Tr}_E\left(\mathrm{Tr}_S\left|E_k\right\rangle\left\langle E_k\right|\mathrm{Tr}_S\left|E_l\right\rangle\left\langle E_l\right|\right)\\
&= \sum_{k\neq l}\mathrm{Tr}_E\left[\mathrm{Tr}_S\left(\left|c_k\right|^2\left|E_k\right\rangle\left\langle E_k\right|\right)\mathrm{Tr}_S\left(\left|c_l\right|^2\left|E_l\right\rangle\left\langle E_l\right|\right)\right]\\
&= \mathrm{Tr}_E\,\omega_E^2 - \sum_k\left|c_k\right|^4\mathrm{Tr}_S\left[\left(\mathrm{Tr}_E\left|E_k\right\rangle\left\langle E_k\right|\right)^2\right]\\
&\leq \mathrm{Tr}_E\,\omega_E^2,
\end{aligned} \tag{1.53}$$

where $\omega_E = \mathrm{Tr}_S\,\omega$. To obtain a further bound, we invoke weak sub-additivity of the Reényi entropy [30]

$$\mathrm{Tr}\left(\omega^2\right) \geq \frac{\mathrm{Tr}_E\left(\omega_E^2\right)}{\mathrm{rank}\left(\rho_S\right)} \geq \frac{\mathrm{Tr}_E\left(\omega_E^2\right)}{d_S}, \tag{1.54}$$

and therefore combining (1.51), (1.53) and (1.54) we get

$$\langle D\left[\rho_S(t),\omega_S\right]\rangle_t \leq \frac{1}{2}\sqrt{d_S\,\mathrm{Tr}_E\left(\omega_E^2\right)} \leq \frac{1}{2}\sqrt{d_S^2\,\mathrm{Tr}\left(\omega^2\right)}, \tag{1.55}$$

which by taking the definition of effective dimension, we get the main result shown in [7]

$$\langle D\left[\rho_S(t),\omega_S\right]\rangle_t \leq \frac{1}{2}\sqrt{\frac{d_S}{d^{\mathrm{eff}}\left(\omega_E\right)}} \leq \frac{1}{2}\sqrt{\frac{d_S^2}{d^{\mathrm{eff}}(\omega)}}. \tag{1.56}$$

As we can see the result obtained by Linden et. al. tell us that the vast majority of quantum systems which the dynamic of the universe is governed by a Hamiltonian with no gaps, will

[26] This condition is reflected in the evaluation by considering the terms where $k \neq \ell$ and $m \neq n$, leading to $m = \ell$ and $k = n$ are the only terms that contribute.

spend most of its time close to its equilibrium state independently of its initial state[27].
To illustrate this result we will picture an universe as the one described before in which the islands refer to the systems regions of the universe such that are atypical and the ocean the region which represent the region where the system equilibrates. As shown in the figure 1.3 we can start from a region in the island (out of the equilibrium) and going out to the ocean (region of equilibrium), as is shown we jump from the island to the ocean so that in average we spend most of the time out of the island. Even though, this results tell us that apparently

Figure 1.3: Illustration of the result from Linden et. al. drawn in a map chart. The ship here refers to the state sailing in our universe. Whereas the islands refer to the region in which our ship is out of equilibrium.

the quantum property of systems is to be close to its equilibrium state, it does not take into consideration how fast it is getting to this state. Since the time averages are taken, we can say that the term of λ goes to zero in trace norm, that is $\langle \|\lambda\|_1 \rangle_t \to 0$, this can be assured since the *Random phase postulate* described above, holds. However, it is easy to see that if we take a state which its energy eigenvalues are very close to each others, the time it will take to equilibrate is exponentially large. It is not of our purpose to measure how fast a

[27] Notice that the last result is not necessarily considering that the state of equilibration will coincide with the canonical state, this result is very general. Exponential bounds can be achieve if we consider the energy eigenvalues of the Hamiltonian to have no rational dependencies, this restriction leads to exponential bound in $d^{\text{eff}}(\omega)$ (check Appendix C in [7]) .

system presenting but instead we present an alternative to the *Random phase postulate*, for the special case in which the energy eigenvalues are very close to each other.
To better explain our idea consider an universe $|\Psi\rangle$ which is decomposed on two states that are very close to each other (meaning that live on the shell of a defined energy), for example a state of the universe belonging to an shell of a defined energy, the energy eigenstates $|E_n\rangle$ and $|E_m\rangle$, the state of the universe can be written as

$$\begin{aligned}\rho &= |\Psi\rangle\langle\Psi| \\ &= \|c_n\|^2 |E_n\rangle\langle E_n| + \|c_m\|^2 |E_m\rangle\langle E_m| \\ &+ c_n c_m^* |E_n\rangle\langle E_m| + c_m c_n^* |E_m\rangle\langle E_n| ,\end{aligned} \tag{1.57}$$

if we take the partial trace of the equation (1.57) this will lead us to the state of the system,

$$\begin{aligned}\rho_S &= \mathrm{Tr}_E\,\rho = \mathrm{Tr}_E\,|\Psi\rangle\langle\Psi| \\ &= \|c_n\|^2 \mathrm{Tr}_E\,|E_n\rangle\langle E_n| + \|c_m\|^2 \mathrm{Tr}_E\,|E_m\rangle\langle E_m| \\ &+ c_n c_m^* \mathrm{Tr}_E\,|E_n\rangle\langle E_m| + c_m c_n^* \mathrm{Tr}_E\,|E_m\rangle\langle E_n| .\end{aligned} \tag{1.58}$$

Something we notice from (1.58) is that since we consider the state to have its eigenvalues of energy to be quite close to each other, by using typicality we get that $\mathrm{Tr}_E\,|E_n\rangle\langle E_n| = \mathrm{Tr}_E\,|E_m\rangle\langle E_m| = \Omega(E)$, with $\Omega(E)$ the canonical state. Therefore the equation (1.58) can be written as

$$\rho_S = \Omega(E) + c_n c_m^* \mathrm{Tr}_E\,|E_n\rangle\langle E_m| + c_m c_n^* \mathrm{Tr}_E\,|E_m\rangle\langle E_n| . \tag{1.59}$$

Since the canonical state is time independent we see that the last terms in (1.59) should cancel in order to be consistent with the result of typicality discussed before. The only way those terms vanish, is by a property that we named call ultra-orthogonality, which refers to an apparent orthogonality of states with respect to the partial trace, that it

$$\mathrm{Tr}_E\,|E_n\rangle\langle E_m| = \mathrm{Tr}_E\,|E_m\rangle\langle E_n| \approx O\left(\sqrt{\frac{d_S}{d^{\mathrm{eff}(\omega_E)}}}\right) \tag{1.60}$$

Considering this, we state that the mechanism which makes the systems to equilibrate emerge naturally as a consequence of typicality.
From these ideas we decided to explore this super-orthogonality in a physical system such that can be solved analytically (or numerically). Take into account that the latter requirement is needed to have a complete knowledge of the state of our universe, and be able to then take a small portion of it to study this reduced state. Thus in order to do that, it is necessary to be able to compute with relative easiness the diagonalization of our universe as well as the reduced states of it in an efficient way[28]. In the following sections we will

[28]We emphasise that it has to be efficient because the trend of exponential growth inherit by Hilbert spaces turns out to be a challenging problem when working with large systems.

provide the background required to understand the choice of the system we made as well as a detailed calculations which provide a full characterization of this system and its reduced states. More over we show a way to generalise this result to any fermionic system and how this Ultra orthogonality is related with a minimal distance code.

Chapter 2

Fermionic States, its characterisation and the connection with Coding theory.

Up to this point we provided a detailed explanation of how in quantum mechanical description, reaching equilibrium emerges as a consequence of the structure of quantum mechanics. Even more, we stated that in the case we deal with states which its eigenenergies are close to each other, typicality will also provide us an answer of how this kind of states will reach equilibrium. As we would like to illustrate this phenomena occur, we focus our study in the fermionic case. In this Chapter we are Going to provide a background to understand how fermionic states are usually treated and why we choose to work with them. More specifically, we provide the overview of solvable fermionic systems, its connection to Majorana fermions and Gaussian states, the formalism of Grassmann for anticommuting variables, and the link between all the formalism for fermions and coding theory.

2.1 Overview

In many areas of physics one has to deal with solving quantum many body problems, which is often a computationally difficult if not impossible task. However, the cases which can be annalitically solved are very well known, and some assumptions have to be taken into account. In spite of this considerations it has been found that a wide class of complicated Hamiltonians with many-body interactions can be often be mapped onto Hamiltonians that are quadratic in annihilation and creation operators and have the generic form [31]

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

where i, j run from 1 to the number of modes in the system (N) and $\hat{a}_i$, $\hat{a}_i^\dagger$ are Fermionic annihilation and creation operators which satisfy the canonical anti-commutation relations (CAR) [32]

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

A convenience when working with these kind of Hamiltonians is that can be diagonalized via a Bogoliubov- Valantin transformations transformation (i.e., canonical transformations), which maps Fermionic creation and annihilation operators on the creation and annihilation operators of non-interacting quasi-particles [33, 34]. Explicitly the transformation looks like

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

where γ_i, κ_i are complex numbers such that preserves the canonical anti-commutation relations given by (2.2) for $\hat{q}$, $\hat{q}^\dagger$[1].

Many relevant physics models are diagonalizable via a Bogoliubov-Valantin transformations, some examples are the Hubbard model, BCS theory of superconductivity in the mean field or Hartree-Fock approximation, and certain solvable spin-chain models (After a Jordan-Wigner transformation) [32]. As we will later explain, an important feature about the class of Hamiltonians described by (2.1), if not the most important for the purpose of this project, is that not only the ground state (quasi-particle vacuum) but all eigenstates describing a excitation in a set of quasi-particles, belong to the class of so-called Fermionic Gaussian states [31]. What is important about this class of states is that are fully characterized by second order correlations, because all the higher moments factorize. This result is very well known and as Wick theorem [35, 36].

2.2 Majorana Fermions

Majorara fermions are fermions such that they are their own antiparticle. In the frame of condensed matter Majorana Fermions (quasi-particles) can be interpreted as a superposition of a electron state and a hole [37].

The formalism of Majorana fermions or Majorana modes of the system (with N modes) can be introduce with the operators

$$\hat{c}_{2j-1} = \hat{a}_j^\dagger + \hat{a}_j, \quad \hat{c}_{2j} = (-i)\left(a_j^\dagger - a_j\right). \tag{2.4}$$

[1] This relation can also be expresses as a condition over γ_i, κ_i,

$$\gamma_i^2 + \kappa_i^2 = 1,$$

and

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

In which case its canonical anti-commutation relations (CAR) take the form

$$\{\hat{c}_k, \hat{c}_l\} = 2\delta_{kl}. \tag{2.5}$$

The anti-commutation relations is seen to be a consequence of an $\mathbb{R}^{2N}$ Clifford algebra[2]. Transformation in between Fermionic and Majorana operators is achieved by matrix of the block form

$$\Omega = \begin{pmatrix} \mathbb{I} & \mathbb{I} \\ i\mathbb{I} & -i\mathbb{I} \end{pmatrix}, \tag{2.6}$$

and then the map from Fermionic ($\vec{\hat{a}}^T = (\hat{a}_1, \ldots, \hat{a}_1^\dagger, \ldots)$) and Majorana ($\vec{\hat{c}}^T = (\hat{c}_1, \ldots, \hat{c}_1^\dagger, \ldots)$) operators is written as $\Omega\vec{\hat{a}} = \vec{\hat{c}}$.
By changing from Fermionic operators to Majorana operators is possible, and convenient, to define de Fermionic covariance matrix which as we mentioned before, fully characterise Gaussians states[3].

2.3 Fermionic Covariance matrix

A system of N fermion modes, described by a set of creation and annihilation operators $\hat{a}^\dagger, \hat{a}$ and satisfying the canonical anti-commutations relations in (2.2), is Gaussian if for such system any state ρ can be written as [38]

$$\rho = \bigotimes_{k=1}^{N} \tilde{\rho}_k, \quad \tilde{\rho}_k = \frac{1}{2}\left(1 - \lambda_i \left[\tilde{a}_i^\dagger, \tilde{a}_i\right]\right), \tag{2.7}$$

for a certain choice of mode basis $\tilde{a} = u_i^j \hat{a}_j + v_i^j \hat{a}_j^\dagger$, and with $|\lambda_i| \leq 1$, where the equality holds for pure states. Equivalently, as mentioned before, Gaussian States are fully characterized by their second moments, so an equivalent form of writing ρ is

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

with $\hat{c} = (\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_{2N})$, the vector of Majorana operators (2.4), Z a normalization constant and G real anti-symmetric $2N \times 2N$ matrix.
Since G is a skew-symmetric matrix, it can always be brought to the block diagonal form

$$OGO^T = \bigoplus_{i=1}^{N} \begin{pmatrix} 0 & -\beta_j \\ \beta_j & 0 \end{pmatrix} \quad \text{with} \quad O \in \mathrm{SO(2N)}, \tag{2.9}$$

[2]By inspection of (2.5) we see that any linear transformation of the form $\tilde{\gamma}_\alpha = O_{\alpha\beta}\gamma_\beta$ where $O \in SO(2N)$, the special orthogonal group in $2N$ dimensions

[3]In comparison to its boson counterpart the fermion Gaussian states have the property that correlation functions for the creation/annihilation operators are completely determined by the two-point functions according to Wick's theorem [35], and moreover, since this property is extensible to correlation function pertaining to a reduced subset of the modes, it follows that any partial (reduced) density matrix obtained from ρ remains Gaussian.

by a special orthogonal matrix $O \in SO(2N)$ where the β_j are called the Williamson eigenvalues of the matrix G. From equation (2.8) it is clear that Gaussian states have an interpretation as thermal (Gibbs) states corresponding to a Hamiltonian of the form

$$\hat{H} = \frac{i}{4}\hat{c}^T G \hat{c} = \frac{i}{4}\sum_{k>l} G_{kl}\left[\hat{c}_k, \hat{c}_l\right] \tag{2.10}$$

and (2.9) shows that every Gaussian state has a normal mode decomposition in terms of N single mode "thermal states" of the form (2.7) ($\sim \exp(-\beta \hat{a}^\dagger \hat{a})$) [39]. From this is clear, that the state can be fully determined by the expectation values of quadratic operators ($\hat{a}_i^{(\dagger)}\hat{a}_j^{(\dagger)}$ and $\hat{a}_i^\dagger \hat{a}_j$). So collecting these expectation values in a real and skew-symmetric covariance matrix Γ which is defined via

$$\Gamma_{kl} = \frac{i}{2}\operatorname{tr}\left(\rho\left[c_k, c_l\right]\right). \tag{2.11}$$

We will be able to bring this anti-symmetric matrix to its block diagonal form, via a canonical transformation.

$$\tilde{\Gamma} = O\Gamma O^T = \bigoplus_{i=1}^{M} \begin{pmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{pmatrix}, \tag{2.12}$$

with $\lambda_i \geq 0$ the Williamson eigenvalues.
In terms of the creation/annihilation operators obtained from the transformed $\tilde{c} = O\hat{c}$, the Gaussian state ρ takes the form (2.8) with $\tilde{\Gamma}$ as its Fermionic covariance matrix. It is easy to see that the relation between G and Γ is given by $\lambda_i = \tanh(\beta_i/2)$, for $i = 1, 2, \ldots, N$ [39].
The equivalence between the special orthogonal group in $2N$ dimensions ($SO(2N)$) and the Fermionic Gaussian states, drives to an interesting property about states describing multi-particles excitations. If $|vacuum\rangle$ is the ground state of some Hamiltonian, with annihilation operators $\hat{a}_i$ in a given quasi-particle basis, then $\hat{a}_i^\dagger |vacuum\rangle = \hat{c}_{2i}|vacuum\rangle$. Meaning that any multi-particle state of this kind is obtained from some transformation (that preserves the canonical anti-commutation relations) of the ground state $|vacuum\rangle$, remains 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 as well as the extension of this property to a reduced subset of the modes, is quite important since a big part of this work is focused on the study of the Fermionic covariance matrix of the XY model.
Up to this point we have talked about some generalities about the quadratic Hamiltonians and how these can be brought to its diagonal form to be analytically solved. However, one may wonder how is this connected to some observables, if the Fermionic covariance matrix has something to do in the observables. The answer to these questions will be boarded in the next section in which we will provide the formalism needed to compute observables over anti-commuting variables as well as the differences we will have between bosons and fermions.

2.4 Grassmann Approach

Fermions are one kind of fundamental particles which can be used to represent quantum information, being one posibility to do quantum computation. From Knill, Laflamme and Milburn [40] work, we know that it is possible to provide an universal set of operations for quantum computation by just using passive linear objects. Right after this discovery, Terhal, DiVincenzo [41] and knill [42] provided a description of the computational capability for Fermionic linear optics (FLO)[4]. These results show that FLO is not a promising way to do quantum computation, since it can be efficiently simulated by classical means, on the other hand, this provides a very interesting tool to study general properties quantum channels for quantum communication.
When working with fermions one could think, weather or not is possible to use the same formalisms than the ones used when working with bosons[5], and as expected many features between them are shared. More specifically, in the case of FLO, it shares most of the features with a limited version of photon linear optics, where mostly Gaussian states appears [40][6].
It is worth mentioning that some of the results in here are quite general and can be used even beyond the scope of our work.

2.4.1 Fermionic Linear Optics

In order provide a more specific description of FLO, we start by defining N abstract Fermionic modes which can be defined by creation $\hat{a}_j$ and annihilation $\hat{a}_j^\dagger$, $j = 1, \ldots, N$, satisfying (2.2). As mentioned above, any state of the system can be generated from the vacuum state $|vacuum\rangle$ of the Fock basis, which we define as

$$|N_1, \ldots, N_N\rangle = \left(\hat{a}_1^\dagger\right)^{N_1} \cdots \left(\hat{a}_n^\dagger\right)^{N_n} |vacuum\rangle \,, \tag{2.13}$$

with $N_i \in \{0, 1\}$. We have then that any sequence $\{N_i\}$ can be understood as a Fermionic state which comes from a superposition of the Fock basis.
When working with FLO, quadratic Hamiltonians govern the dynamics, in which terms associated with individual energy modes, tunneling and bulk dynamics are included. The structure of these Hamiltonians is what allow us conveniently work with operators which generate the *Clifford* algebra ($\mathcal{C}_{2N}$), described in (2.5). So we will have that any arbitrary

[4]Whenever we are referring to this term, it basically has in consideration a system consisting of non-interacting electrons in a controllable external potential and a detector that measures projectively occupation numbers of single-electron modes.

[5]By this we mean that it could be also written in its correspond operators of annihilation and creation $\hat{a}^\dagger$,$\hat{a}$ [43]

[6]the main reason for this is that a set of states that can be achieved by FLO operations starting from the Fock vacuum is a set of Fermionic Gaussian states, just as the case of its bosonic counterpart.

operator $X \in \mathcal{C}_{2N}$ will be written as a polynomial in the operators $\{\hat{c}_a\}$ as

$$X = \alpha \hat{I} + \sum_{p=1}^{2n} \sum_{1 \leq a_1 < \cdots < a_p \leq 2n} \alpha_{a_1,\ldots,a_p} \hat{c}_{a_1} \cdots \hat{c}_{a_p}, \tag{2.14}$$

with $\alpha = 2^{-n} \operatorname{Tr}(X)$. Particularly for the Hamiltonian operator we have that

$$H = \frac{i}{4} \sum_{a,b=1}^{2N} H_{ab} \hat{c}_a \hat{c}_b, \tag{2.15}$$

with $\{H_{ab}\}$ an anti-symmetric $2N \times 2N$ matrix.
We may think that when working with fermions could be the same as working with bosons by changing commutators with anti-commutators and changing to a fine algebra. However, this is not the case, as we will show in the next section observables have to be computed by using the corresponding calculus for fermionic systems.

2.4.2 Grassmann Calculus

There are a lot of results from the earlier seventies which describe techniques to study systems with infinite number of fermionic modes [44], nonetheless, the formalism we are going to present here is a specific case where a finite number of modes is considered, and consist in a customization of the Lagrangian representation for infinite number of Fermionic modes, presented in [45]. Consider $\theta_1, \ldots, \theta_N$ span a N-dimensional complex linear space $\mathbb{C}^N$. From the anti-commutation relation we get that for $\theta_1, \ldots, \theta_N$,

$$\theta_a^2 = 0 \quad \text{and} \quad \theta_a \theta_b + \theta_b \theta_a = 0, \tag{2.16}$$

so it is clear that the most general function we can built over the Grassmann algebra with complex coefficients $\mathcal{G}_N$ is a polynomial of θ's

$$f(\theta) = \alpha + \sum_{p=1}^{n} \sum_{1 \leq a_1 < \cdots < a_p \leq N} \alpha_{a_1,\ldots,a_p} \theta_{a_1} \cdots \theta_{a_p}, \tag{2.17}$$

where the coefficients α_* are complex numbers. A polynomial $f(\theta)$ will be called *even* if it involves only even powers of θ, nonetheless something to stress is that even elements constitute the center of the Grassmann algebra.
It is possible to differentiate functions of Grassmann variables. A partial derivative over θ_a is a linear operator from $\mathcal{G}_N$ to $\mathcal{G}_N$, which is defined by

$$\frac{\partial}{\partial \theta_a} 1 = 0, \quad \frac{\partial}{\partial \theta_a} \theta_b = \delta_{ab}, \tag{2.18}$$

and Liebniz's rule

$$\frac{\partial}{\partial \theta_a} \left(\theta_b f(\theta) \right) = \delta_{ab} f(\theta) - \theta_b \frac{\partial}{\partial \theta_a} f(\theta). \tag{2.19}$$

The equality (2.19) implies that

$$\left\{ \frac{\partial}{\partial\theta_a}, \frac{\partial}{\partial\theta_b} \right\} = 0. \tag{2.20}$$

Since a derivative $\frac{\partial}{\partial\theta_a} f(\theta)$ does not depend upon variable θ_a, it is sometimes convenient to think about differentiation as a linear operator mapping $\mathcal{G}_n$ into $\mathcal{G}_{n-1}$. Such operator is called integration and is denoted as [46] [7]

$$\int d\theta_a \equiv \frac{\partial}{\partial\theta_a} : \mathcal{G}_n \to \mathcal{G}_{n-1}. \tag{2.21}$$

From equation (2.20) we have that

$$\int \mathrm{D}\theta \frac{\partial}{\partial\theta_a} f(\theta) = 0, \tag{2.22}$$

this equation combined with (2.19) give us the anti-commuting version of integration by parts.
With these properties we show the main equation that will help us calculate what is need for the purpose of this work.
For a vectors of Grassmann variables $\vec{\theta}$, $\vec{\eta}$ and complex anti-symmetric matrix M we have

$$\int \mathrm{D}\theta \exp\left(\frac{i}{2}\theta^T M\theta\right) = i^n \operatorname{Pf}(M), \tag{2.23}$$

and

$$\int \mathrm{D}\theta \exp\left(\eta^T\theta + \frac{i}{2}\theta^T M\theta\right) = i^n \operatorname{Pf}(M) \cdot \exp\left(-\frac{i}{2}\eta^T M^{-1}\eta\right). \tag{2.24}$$

In these formulas $\operatorname{Pf}(N)$ is the Pfaffian of a complex antisymmetric matrix N defined as anti-symmetrized product $\mathcal{A}(N_{1,2}N_{3,4}\cdots N_{2n-1,2n})$ that is

$$\operatorname{Pf}(N) = \frac{1}{2^n n!} \sum_{\sigma\in S_{2n}} \operatorname{sgn}(\sigma) N_{\sigma_1,\sigma_2}\cdots N_{\sigma_{2n-1}\sigma_{2n}} \tag{2.25}$$

2.4.3 Gaussian States

For this part we are going to introduce in a formal way the states we have been talking about in at the beginning of this chapter. Informally we can say that any Gaussian operator can be

[7] Here we will use a compact notation

$$\int \mathrm{D}\theta \equiv \int d\theta_n \cdots \int d\theta_2 \int d\theta_1,$$

and the order is chosen such that

$$\int D\theta\ \theta_1\cdots\theta_n = 1$$

.

represented as an exponent of another operator which is quadratic in creation/annihilation operators. When considering an operator $X \in \mathcal{C}_{2N}$ It is natural to assign a polynomial $\omega(X,\theta) \in \mathcal{G}_{2N}$ of $2N$ Grassmann variables defined by

$$\omega\left(\hat{c}_p\hat{c}_q\cdots\hat{c}_r,\theta\right) = \theta_p\theta_q\cdots\theta_r, \quad \omega(\hat{\mathbb{I}},\theta) = 1, \tag{2.26}$$

where this definition comes from the general functions in the Grassmann algebra described in (2.17).
To illustrate what how this map can be done, consider as an example the projector $\hat{a}_1\hat{a}_1^\dagger$, which its action is nothing but to project into a state where the first mode is empy. For this precise case we have that the map to Grassmann variables will look as

$$\omega(\hat{a}_1\hat{a}_1^\dagger,\theta) = \frac{1}{2}\left(\mathbb{I} + i\theta_1\theta_2\right) = \frac{1}{2}\exp\left(i\theta_1\theta_2\right), \tag{2.27}$$

where all exponents were defined by a Taylor series.

For any two operators $X, Y \in \mathcal{C}_{2N}$ one can compute the trace by just using a simple formula,

$$\mathrm{Tr}(XY) = (-2)^N \int D\theta D\mu e^{\theta^T\mu}\omega(X,\theta)\omega(Y,\mu), \tag{2.28}$$

which can be directly verified. It is also easy to see that canonical transformations of an operator $X \in \mathcal{C}_{2N}$ are equivalent to an orthogonal change of basis in the space of the Grassmann variables, this is

$$\omega\left(VXV^\dagger,\theta\right) = \omega(X,\eta), \quad \eta_a = \sum_{b=1}^{2n} R_{ab}\theta_b. \tag{2.29}$$

So we will define a Gaussian states of N Fermionic modes as via its density operator $\rho \in \mathcal{C}_{2N}$. We say that a state is Gaussian in $\mathcal{C}_{2N}$ *iff* its Grassmann representation is as well Gaussian

$$\omega(\rho,\theta) = \frac{1}{2^n}\exp\left(\frac{i}{2}\theta^T M\theta\right), \tag{2.30}$$

for some $2N \times 2N$ antisymmetric matrix M. The matrix M is defined as

$$M_{ab} = \frac{i}{2}\mathrm{Tr}\left(\rho\left[\hat{c}_a,\hat{c}_b\right]\right) = \begin{cases} \mathrm{Tr}\left(\rho i\hat{c}_a\hat{c}_b\right) \text{ for } a \neq b \\ 0 \text{ for } a = b \end{cases}, \tag{2.31}$$

which is nothing but the covariance Fermionic matrix defined in (2.11), meaning that it can be brought to its Williamson form to get its eigenvalues.

Therefore the connection between the Grassmann formalism for Gaussian states and the Fermionic covariance matrix described at the beginning of the chapter, is easily understood

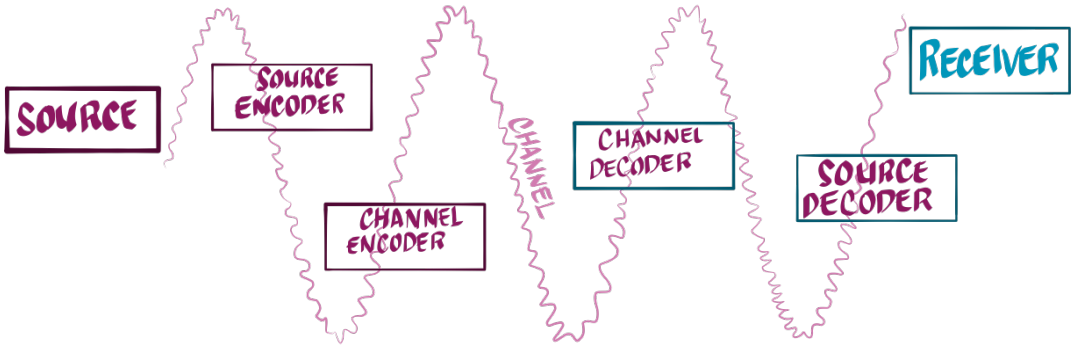


Figure 2.1: Representation of the scheme of communication. In the image the noise in the channel is represented by the noise wavy connection between the parts in the communication.

by simply assigning the correspondent covariant matrix to the states we work with. Nonetheless, one could be wondering how is this formalism connect to the error correcting code theory. In the next section we will provide a little historical background of how this theory was done and how one can think states of fermions as binary codes. More specifically we will provide the necessary background to link the theory of correcting errors with Fermionic states which are constrained over a shell of energy in the Hilbert Space.

2.5 Error correcting Code Theory

In this section we will provide some concepts and definitions in order to understand how the Coding theory can be linked to Fermionic states.
In 1948, Claude Shannon presents his extraordinary work named "A Mathematical Theory of Communication" [24] in which he provided a precise measure of the information content of a random variable in terms of its entropy. His work is divided in two parts the noiseless coding theorem and the noisy channel theorem. For the purpose of our needs we are going to focus only in the second part of his work, which states that a reliable communication is possible if we use schemes such that its rate is less than the capacity of the channel. Even though he never provided an idea of how this schemes could be found, his work is considered one of the most relevant discovery of the century.

Here we will consider codes in communication scenario, as the one showed in figure 2.1,

meaning that there will be a sender who wants to send k message symbols over a noisy channel and there will be a receiver who has to correct possible errors over the sent code to fully interpret it. The sender will first encodes the k message symbols into n symbols. The receiver then tries to recover the original k message symbols. thus, encoding is the process of adding redundancy and decoding is the process of removing errors and the communication can only be done over the channel [23]. The most fundamental question one can ask is what will be the relation between the amount of redundancy and the errors that can be corrected, and in order to answer this question we will provide some useful definitions.

2.6 Some basic definitions

Definition 2.6.1 (Code)**.** *A code of block C length n over an alphabet Σ is a subset of Σ^n. If $|\Sigma| = q$, we say that C is a q-ary code.*

it is worth mention that, associated with a code there is also an encoding map E which maps the message set $\mathcal{M}$, identified in some canonical way with $\{1, 2, \ldots, |C|\}$ say, to code words belonging to Σ^N, and thus we have to understand the code as the image of the encoding map [47].

Definition 2.6.2 (Dimension of a code)**.** *Given a code $C \subset \Sigma^n$, its dimension is given by*

$$k \stackrel{\text{def}}{=} \log_q |C|, \tag{2.32}$$

An interesting fact about defining the dimension of the code in this way is that implicitly it is telling us that when working with codes exponential growth will be always taken into account.
We have to provide here a way to measure the amount of redundancy in a given message.

Definition 2.6.3 (Rate of a code)**.** *The rate of a code with dimension k and block length n is given by*

$$R \stackrel{def}{=} \frac{k}{n}, \tag{2.33}$$

this definition is nothing but the average amount of non redundant information each of the n symbols transmitted over the channel.
However, an alternative, and more general way of defining this is via the size of the code and the alphabet as

$$R(C) = \frac{\log |C|}{n \log |\Sigma|}. \tag{2.34}$$

Definition 2.6.4 (Hamming distance)**.** *The Hamming distance between two strings x and y of the same length over a finite alphabet Σ, denoted $\Delta(x, y)$, is defined as the number of positions at which the two strings differ, i.e, $\Delta(x, y) = |\{i | x_i \neq y_i\}|$. The fractional Hamming distance or relative distance between $x, y \in \Sigma^n$ is given by $\delta(x, y) = \frac{\Delta(x,y)}{n}$.*

It is trivial to check that the Hamming distance defines a metric on Σ^n.

Definition 2.6.5. ***Hamming weight:*** *The Hamming weight of a string x over alphabet Σ is defined as the number of non-zero symbols in the string. More formally, the Hamming weight of a string $\mathcal{W}(x) = |\{i|x_i \neq 0\}|$. Note that $\mathcal{W}(x-y) = \Delta(x,y)$.*

Given a string $x \in \Sigma^N$, the Hamming ball of radius r around x is the set $\{y \in \Delta^n | \Delta(x,y) \leq r\}$.
The minimum distance, or simply distance, of a code C, denoted $\Delta(C)$, is defined to be the minimum Hamming distance between two distinct code words of C. That is

Definition 2.6.6 (Minimum distance)**.** *The minimum distance, or simply distance, of a code C, denoted $\Delta(C)$, is defined to be the minimum Hamming distance between two distinct code words of C. That is*

$$\Delta(C) = \min_{\substack{c_1,c_2 \in C \\ c_1 \neq c_2}} \Delta(c_1, c_2). \tag{2.35}$$

In particular, for every pair of distinct code words in C the Hamming Distance between them is at least $\Delta(C)$

The relative distance of C, denoted $\delta(C)$, is the quantity $\frac{\Delta(C)}{N}$, where N is the block length of C. Thus any two code words of C differ in at least a fraction $\Delta(C)$.

Definition 2.6.7 (Notation)**.** *A q-ary code of block length N and dimension k will be referred to as an $[N,k]_q$ code. Further, if the code has minimum distance d, it will be referred to as an $[N,k,d]_q$ code. When the alphabet size q is clear from the context, or not very relevant to the discussion, we omit the subscript.*

Up to this point we have only described specific codes, codes with fixed block length and dimension. However, since we are interested in the asymptotic behaviour, it turns out to be more useful the study of families of codes instead of an specific code.

Definition 2.6.8 (Family of Codes)**.** *Let $q \geq 2$. let $\{n_i\}_{i\geq 1}$ be and increasing sequence of block lengths and suppose there exists sequences $\{k_i\}_{i\geq 1}$ and $\{d_i\}_{i\geq 1}$ such that for all $i \geq 1$ there exist an $[n_i, k_i, d_i]_q$ code C_i. then the sequence $C = \{C_i\}_{i\geq 1}$ is a family of codes. The rate of C is defined as*

$$R(C) = \lim_{i\to\infty} \left\{ \frac{k_i}{n_i} \right\}, \tag{2.36}$$

and the relative distance of C is defined as

$$\delta(C) = \lim_{i\to\infty} \left\{ \frac{d_i}{n_i} \right\}, \tag{2.37}$$

from now on whenever we talk about a code we are implicitly referring to a family of codes.

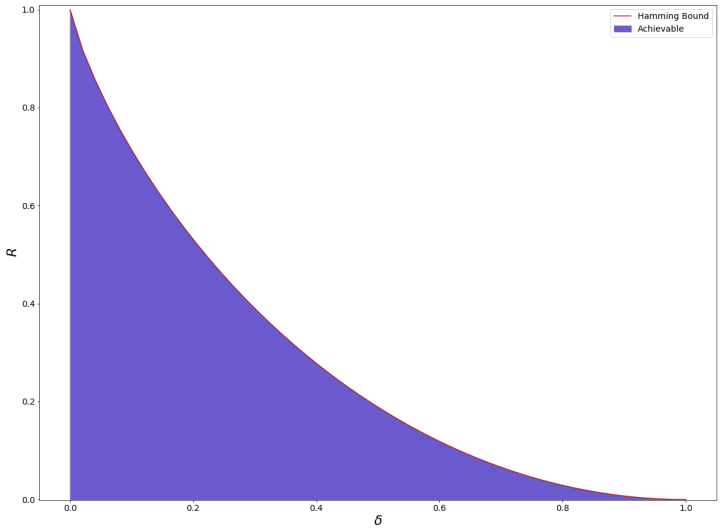


Figure 2.2: An illustration of the Hamming bound for the case of $q = 2$. Note any code above this bound could exist. The shaded region shows the codes that could exist.

For the purpose of this work, we are going to focus on a particular class of codes, name minimum distance code. This special kind of codes came to our interest since they seem to appear naturally when study typical Fermionic states. We will then show some of the main results on this particular kind of codes to after show how this results could be extended to our particular case.
When we talk about codes of minimum distance d we refer to codes which have the property that for every pair x, y of codewords we have that

$$\Delta(x, y) = d, \tag{2.38}$$

and therefore, the its relative distance δ is given by $N\delta = d$.
when working with this kind of codes one may wonder what is the best rate we can achieve. Particularly we are going to show 3 results, 1 positive and 2 negative results[8]. Even though there are other known bounds, we will not talk about others but Gilbert-Varshamov, Hamming and Plotkin bounds, the reason for this is due to the fact that the other bound apply for large enough alphabets, so for binary codes we are not interested at all in these kind of bounds [23].

[8]Note that a negative result refer to an upper bound on the rate, meaning that the maximum achievable rate we could get can not exceed some value. Whereas a positive result refer to lower bound on the rates we can achieve.

2.6.1 Hamming Bound (Sphere packing bound)

Definition 2.6.9 (Volume of Hamming ball)**.** *Let $q \geq 2$ and $n \geq r \geq 1$ be integers. Then the volume of a Hamming ball of radius r is given by*

$$\mathrm{Vol}_q(r,n) = |B_q(\mathbf{0},r)| = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i, \tag{2.39}$$

where the choice of $\mathbf{0}$ as the center of the Hamming ball is chosen arbitrary, since the volume of the Hamming ball is independent of its center.

It is simple to show that

$$\frac{k}{n} \leq 1 - \frac{\log_q \mathrm{Vol}_q\left(\left\lfloor \frac{d-1}{2} \right\rfloor, n\right)}{n}, \tag{2.40}$$

where the volume in (2.40) correspond to the definition in(2.39). With some algebra and using the Stirling asymptotic approximation one can show that

$$\mathrm{Vol}_q\left(\left\lfloor \frac{d-1}{2} \right\rfloor, n\right) \geq q^{H_q\left(\frac{\delta}{2}\right)n - o(n)}, \tag{2.41}$$

where the latter inequality immediately provide us an upper bound on the rate

$$R \leq 1 - H_q\left(\frac{\delta}{2}\right) + o(1). \tag{2.42}$$

The inequality in (2.42) is known as the Hamming Bound.

2.6.2 Plotkin Bound

Definition 2.6.10 (Plotkin Bound)**.** *The following holds for any code $\mathcal{C} \subset [q]^n$*

- *If $d = \left(1 - \frac{1}{q}\right) n, |\mathcal{C}| \leq 2qn$.*
- *If $d > \left(1 - \frac{1}{q}\right) n, |\mathcal{C}| \leq \frac{qd}{qd-(q-1)n}$.*

Note that the Plotkin Bound implies that a code with relative distance $\delta \geq 1 - 1/q$, must necessarily have $R = 0$.

Definition 2.6.11. *For any q-ary code with relative distance $0 \leq \delta \leq 1 - \frac{1}{q}$,*

$$R \leq 1 - \left(\frac{q}{q-1}\right)\delta + o(1). \tag{2.43}$$

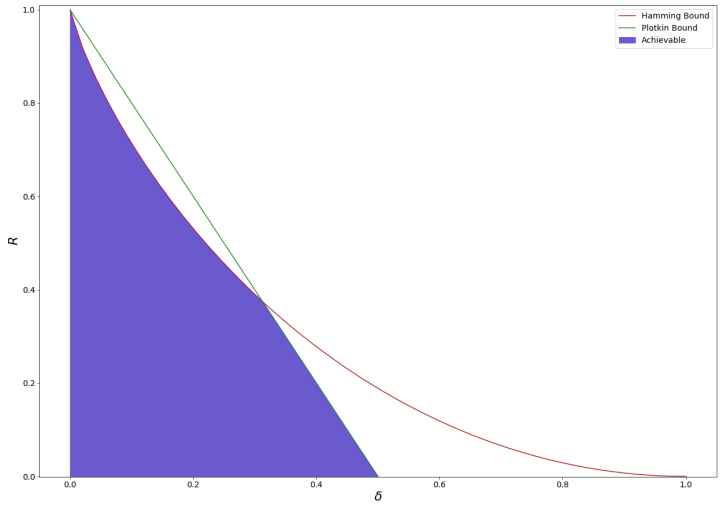


Figure 2.3: An illustration of Plotkin and Hamming bounds for the case of $q = 2$. For this case the shaded region changes indicating us that there are not codes with $R > 0$ when $\delta = 1 - 1/q$.

To illustrate the proof of this bound we can consider the distance $d = n\delta$. So we can shorten the codewords and group them in a way such that they agree on the first $n - n'$ symbols, with $n' = \left\lfloor \frac{qd}{q-1} \right\rfloor - 1$. Then in particular for any $x \in [q]^{n-n'}$, define the prefix code

$$\mathcal{C}_{\mathbf{x}} = \{(c_{n-n'+1}, \dots c_n) \mid (c_1 \dots c_N) \in \mathcal{C}, (c_1 \dots c_{n-n'}) = \mathbf{x}\} \,. \tag{2.44}$$

for all x, $\mathcal{C}_x$, has distance d as $\mathcal{C}$ has distance d. Additionally, it has block length $n' < \left(\frac{q}{q-1}\right) d$, and thus $d > \left(1 - \frac{1}{q}\right) n'$. From the Plotkin bound, this implies that

$$|\mathcal{C}_{\mathbf{x}}| \leq \frac{qd}{qd - (q-1)n'} \leq qd, \tag{2.45}$$

where the second inequality follows from the fact that $qd - (q-1)n'$ is an integer.
Note that from the definition of $\mathcal{C}_x$

$$|\mathcal{C}| = \sum_{\mathbf{x} \in [q]^{n-n'}} |\mathcal{C}_{\mathbf{x}}| \,, \tag{2.46}$$

which tell us that

$$|\mathcal{C}| \leq \sum_{\mathbf{x} \in [q]^{n-n'}} qd = q^{n-n'} \cdot qd \leq q^{n - \frac{q}{q-1} d + o(n)} = q^{n\left(1 - \delta \cdot \frac{q}{q-1} + o(1)\right)} \,, \tag{2.47}$$

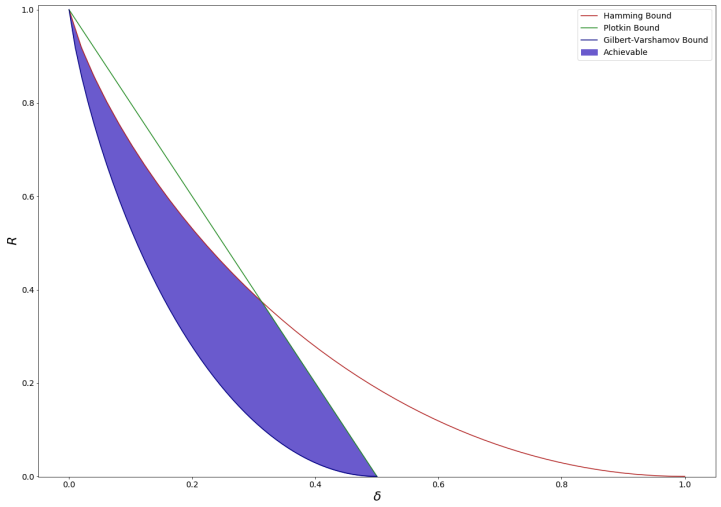


Figure 2.4: An illustration of 3 bounds Hamming, Plotkin and Gilbert-Varshamov for the case of $q = 2$. The lower bound correspond to the Gilvert-Varshamov Bound whereas the other 2 are the upper bound for the rate.

in other words this provides another upper bound to the rate given by

$$R \leq 1 - \left(\frac{q}{q-1}\right)\delta + o(1) \tag{2.48}$$

To close this part we will show the latter but positive result which provide us lower bound on the code rates, Gilbert Varshamov bound.

2.6.3 Gilbert Varshamov Bound

We now switch gears to provide a positive result. We will only provide the main ideas for the proof of this result an we will discuss why this result turn out to be one of the most important results.

Definition 2.6.12 (Gilbert-Varshamov Bound)**.** *Let $q \geq 2$. For every $0 \leq \delta < 1 - \frac{1}{q}$ and $0 < \varepsilon \leq 1 - H_q(\delta)$. There exist a code with rate $R \geq 1 - H_q(\delta) - \varepsilon$ and relative distance δ.*

To provide a main idea about the proof we can consider a greedy approach. First we start with an empty code $\mathcal{C}$ and we keep adding vectors that are not in $\mathcal{C}$ and that have Hamming distance at least d from all the existing codewords in $\mathcal{C}$. Notice that by doing so we can assure that we will never add a vector c that will make that will make the distance of $\mathcal{C}$ fall below d. Indeed it is easy to see that after doing so, we have

$$\bigcup_{\mathbf{c} \in C} B(\mathbf{c}, d-1) = [q]^n, \tag{2.49}$$

this is easily checked, because if it was not true, then there would exist a vector $\mathbf{v} \in [q]^n \backslash C$, such that $\Delta(\mathbf{v}, \mathbf{c}) \geq d$ and therefore $\mathbf{v}$ can be added. How ever, this would contradict the fact that we have finished the procedure. So

$$\bigcup_{c \in C} B(\mathbf{c}, d-1) \mid = q^n. \tag{2.50}$$

It isn't hard to see that

$$\sum_{\mathbf{c} \in C} |B(\mathbf{c}, d-1)| \geq \left| \bigcup_{\mathbf{c} \in C} B(\mathbf{c}, d-1) \right|, \tag{2.51}$$

which implies that

$$\sum_{\mathbf{c} \in C} |B(\mathbf{c}, d-1)| \geq q^n, \tag{2.52}$$

but as mentioned before, the volume of the Hamming ball is translation invariant,

$$\sum_{\mathbf{c} \in C} \mathrm{Vol}_q(d-1, n) \geq q^n. \tag{2.53}$$

Since $\sum_{\mathbf{c} \in C} Vol_q(d-1, n) = Vol_q(d-1, n) \cdot |C|$

$$\begin{aligned} |C| &\geq \frac{q^n}{\mathrm{Vol}_q(d-1, n)} \\ &\geq \frac{q^n}{q^{nH_q(\delta)}} \\ &= q^{n(1-H_q(\delta))}. \end{aligned} \tag{2.54}$$

And therefore concluding the proof. It is worth mention that this way of proceeding the code have not any special structure but as one might think this algorithm will take exponentially long time to finish. How ever, one may wonder if there is a special kind of code that also achieve this rate. Indeed is possible to show that random linear codes lies, with high probability, on the Gilbert-Varshamov Bound. To pick a linear random code we only need pick a random $k \times n$ matrix, in which each entry is chosen uniformly and independently at random according to its alphabet [23,47]. Aside from these result providing some bounds on the rate of the minimum distance codes. We are interested in the performance of a special kind of random codes, more specifically binary codes over a binary-symmetric channel (BSC). In here we derive the minimum distance, distance distribution and error exponent of a typical random code (TRC) from a random code ensemble (RCE), as well as the one correspondent to a typical linear code (TLC) from a linear code ensemble (LCE) [48]. As mentioned by *A Barg, and G. D. Forney, Jr* most of the important of these results are expressed in terms of the Gilbert-Varshamov distance $\delta_{GV}(R)$.

2.6.4 Error Exponents for Random Minimum Distance Codes

It is very well known that on a BSC with crossover probability p, the channel capacity is $C = 1 - H_2(p)$ [9]. The error coding exponent $E_r(R)$ is positive for $0 \leq R < C$ and given by [49, 50].

$$E_r(R) = \begin{cases} R_0 - R, & 0 \leq R \leq R_{\text{crit}} \\ E_{\text{sp}}(R), & R_{\text{crit}} \leq R \leq C \end{cases}, \tag{2.55}$$

where R_0, R_{crit} and $E_{\text{sp}}(R)$, are known as cutoff rate, critical rate and the sphere-packing exponent respectively. In [51] Gallager has shown that the random coding exponent is the true error exponent for the RCE on any discrete memoryless channel. Here we will show the main results provided in [48] and we will provide the ideas to derive these results, as we will show later this ideas of error exponents will be quite helpful to understand how it is possible to make a connection between minimum distance codes and typical Fermionic states.

Random Binary Codes

Consider a binary code C of length N and rate R bits per symbol is a set of $M = 2^{NR}$. For the case of RCE one compute the probability that by taking a random codeword $\mathbf{x}_i$ of length N it would have Hamming distance $d = N\delta$ from an arbitrary binary $N-$tuple $\mathbf{b}$ and see that it will be independent of $\mathbf{b}$ and equals to

$$\Pr\{d_H(\mathbf{x}_i, \mathbf{b}) = d\} = \binom{N}{d} \tilde{p}^d (1-\tilde{p})^{N-d}, \tag{2.56}$$

where $\tilde{p}$ corresponds to the probability of having a one. Under this RCE, two distances $d_H(\mathbf{x}_i, \mathbf{x}_j)$ and $d_H(\mathbf{x}_{i'}, \mathbf{x}_{j'})$ are independent random variables. So if we consider the number of unordered pairs of codewords $(\mathbf{x}_i, \mathbf{x}_j)$ with $i \neq j$ in C at a distance d apart

$$S_{\mathcal{C}}(d) = \sum_{i=0}^{M-1} \sum_{j=0}^{i-1} \Phi\left\{d_H\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = d\right\}, \tag{2.57}$$

where $\Phi\left\{d_H\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = d\right\}$ is equal to 1 if the condition $d_H(\mathbf{x}_i, \mathbf{x}_j)$ is satisfied and 0 otherwise. For the case of RCE on a BSC $S_{\mathcal{C}}(d)$ is a sum of $\binom{M}{2}$ pairwise independent, identically distributed random variables, so we have

$$\mathrm{E}S_{\mathcal{C}}(d) = \left(\begin{array}{c} M \\ 2 \end{array} \right) \mathrm{E}\Phi \doteq 2^{N(2R-1+\mathcal{H}(\delta))}. \tag{2.58}$$

Therefore we are ready to state the following theorem

Theorem 2.6.1 (Minimum distance 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*

[9] We will be using the notation $\mathcal{H}$ to refer to the binary entropy $\mathcal{H} \equiv H_2$

less than $\delta_{GV}(2R) - \varepsilon$ *goes to zero exponentially as* $N \to \infty$. *For* $0 \leq R < 1$, *if* $d = N\delta$ *is such that*

$$\delta_{GV}(2R) + \varepsilon \leq \delta \leq 1 - \delta GV(2R) - \varepsilon, \tag{2.59}$$

then the probability that the number of codeword pairs at a distance d *satisfies* $S_{\mathcal{C}}(d) \doteq 2^{N(2R-1+\mathcal{H}(\delta))}$ *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

$$\Pr\{S_{\mathcal{C}}(d) \geq 1\} \leq \mathrm{E}S_{\mathcal{C}}(d) \doteq 2^{-N(1-\mathcal{H}(\delta)-2R)} \to 0, \tag{2.60}$$

which in other words it tell us that with probability differing from 1 by an exponentially falling quantity, there will be no pairs at distance d. Notwithstanding this result, 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 $\mathrm{E}S_{\mathcal{C}}(d)$ at a distance d is exponentially large. To see this, we can use the Chebyshev inequality, so for any $\delta > 0$, we have

$$\Pr\left\{|S_{\mathcal{C}}(d) - \mathrm{E}S_{\mathcal{C}}(d)| \geq \binom{M}{2}\alpha\right\} \leq \frac{\mathrm{E}\Phi}{\binom{M}{2}\alpha^2}, \tag{2.61}$$

by choosing $\alpha \doteq 2^{-N(1-\mathcal{H}(\delta)+\Delta)} < \mathrm{E}\Phi$ for any $\Delta > 0$, we have

$$\Pr\left\{|S_{\mathcal{C}}(d) - \mathrm{E}S_{\mathcal{C}}(d)| > \binom{M}{2}\alpha\right\} \leq \frac{2\mathrm{E}\Phi}{M(M-1)\alpha^2} \doteq 2^{-N(2R-1+\mathcal{H}(\delta)-2\Delta)}. \tag{2.62}$$

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

Random Linear Codes

A binary linear code C of length N and rate K/N is a set of $M = 2^K$ binary N-tuples that is generated by K $N-$tuples $\mathbf{g}_j$, $1 \leq k \leq K$. This is

$$\mathbf{x}(u) = \Sigma_k \mathbf{u}_k \mathbf{g}_j, \tag{2.63}$$

where $\mathbf{u}$ is an arbitrary binary $K-$tuple. Here we will be considering the case in which each of the 2^{NK} matrices are chosen with equal probability. For the case of linear codes, the distribution of distance $\{d_H(\mathbf{x}_i, \mathbf{x}_j), i \neq j\}$ from any given codeword $\mathbf{x}_i$ is independent of i. The average distance distribution of a linear code C therefore reduces to

$$\mathcal{N}_{\mathcal{C}}(d) = \sum_{j \neq i} \Phi\left\{d_H\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = d\right\} \quad (d = 1, 2, \ldots, N), \tag{2.64}$$

where $\mathbf{x}_i$ is an arbitrary codeword. Typically, $\mathbf{x}_i$ is taken as the all zero codeword $\mathbf{0} = \mathbf{x}(\mathbf{0})$. If $(\mathbf{u}_j, \mathbf{u}_k)$ is any pair distinct non-zero $K-$tuples, then the corresponding codewords $(\mathbf{x}(\mathbf{u}_j), \mathbf{x}(\mathbf{u}_k))$ are a pair of independent random binary $N-$tuples. It follows that two distinct distances $d_H(\mathbf{x}_i, \mathbf{x}_j)$ and $d_H(\mathbf{x}_i, \mathbf{x}_k)$ from a given codeword $\mathbf{x}_i$ are pairwise-independent and distributed as in the RCE. In particular

$$\Pr\left\{d_H\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = N\delta\right\} \doteq 2^{-N(1-\mathcal{H}(\delta))}, \tag{2.65}$$

for any $d = N\delta$, the quantity $\mathcal{N}_{\mathcal{C}}(d)$ in (2.64) is a sum of $M-1 \doteq 2^{NR}$ pairwise-independent, identically distributed random variables with mean $E\Phi \doteq 2^{-N(1-mathcalH(\delta))}$. Its mean value is therefore equal to

$$\mathcal{N}_{\mathrm{LCE}}(d) \doteq 2^{N(R-1+\mathcal{H}(\delta))}. \tag{2.66}$$

Therefore we say that the relative minimum distance of a code chosen at random from the LCE will be, with probability $1 - 2^{-\Omega(N)}$, approximately equal to the GV relative distance $\delta_{GV}(R)$.
In summary, the typical minimum distance in the LCE is better than that in the RCE because the minimum of only $M-1$ pairwise-independent distances, whereas in the RCE it is the minimum of $\binom{M}{2}$ pairwise independent distances.
Up to this point one may be wonder how all this theory of correcting errors and minimum distance codes are connected to our problem. In following section we are going to show how this connection emerge as a natural consequence of the structure in the Clifford algebra and therefore provide an explanation of what we named as "Super-Orthogonality".

2.7 Mechanism behind typicality as a correction error method.

As we mentioned in the first chapter, we are interested in study the behaviour of states which are close in energy. First we will show why it is possible to study "Ultra Orthogonality" for Fermionic states in two possibilities. The first one is in terms of minimum distance codes and we will see that "Ultra Orthogonality" turns out to be an exact result. For the second part, we deduce its correspondent exponent error. For the last part, we will tackle the problem when we can not work over a minimum distance and we will show how "Ultra orthogonality" can be acting over this specific case. Even though for this case we were not able to generalised this to every Fermionic system, we will discuss why this result should holds in general.

2.7.1 "Ultra Orthogonality" on Fermions

To start this part we want to recapitulate a couple of things. First, as we showed in equation (1.49), when we take the partial trace the remaining cross terms will be of our interest. When we discuss the property of Ultra Orthogonality, we stated that the cross terms of $\mathrm{Tr}_E |E_n\rangle \langle E_m|$ should be the ones that had to be near to zero, in order for thermalisation to occur on the specific case we discussed in chapter 1. Here we are going to formalise all these ideas.

We can start by choosing an arbitrary state such that it can be decomposed in its vectors of the Fock basis as[10]

$$|\psi\rangle = \sum_{\vec{n}} \psi(\vec{n}) |\vec{n}\rangle . \tag{2.67}$$

Therefore we are interested in studying the terms of the form[11]

$$\hat{X}_{ij} \doteq \mathrm{Tr}_E \left(|\vec{n}_i\rangle \langle \vec{n}_j| \right) \equiv \mathrm{Tr}_{N/L} \left(|\vec{n}_i\rangle \langle \vec{n}_j| \right) . \tag{2.68}$$

What Ultra orthogonality tell us is that this term has to zero or in the worst scenario something of the order of fluctuations. Since we are working on the very special case of Quadratic Hamiltonian describing Fermionic systems, we recall the fact that the operators are generated by the Majorana operators, and form the so called Clifford algebra, described in section 2.2. We also saw that these operators can be mapped to the Grassman variables, which allow us to compute things like observables. Taking into account that we started with an space in which we had to deal with N Pauli operator and we changed to a new space in which we work with $2N$ spinless operators and the fact that the most general function we can built over the Grassman algebra is polynomial of the Grassmann variables. Thus we construct a function of the Grassman variables which takes two binary sequences $(\vec{x}, \vec{y})$, $x_i, y_i \in \{0, 1\}$, lets call it $\gamma(\vec{x}, \vec{y})$. the reason for defining this function is that we are going to describe the system in the space of size L with its correspondent operators, meaning that these sequences can not be arbitrary, they must have to be sequences such that after its first L elements, they must have only zeros. To illustrate this consider the following example, let $\vec{x} = (0100 \ldots 0)$, $\vec{y} = (1100 \ldots 0)$, here $L = 3$, so our function will be described by[12]

$$\gamma(\vec{x}, \vec{y}) = \gamma_1^{x_1=0} \gamma_4^{y_1=1} \gamma_2^{x_2=1} \gamma_5^{y_2=1} \gamma_3^{x_3=0} \gamma_6^{y_3=1} = \gamma_4 \gamma_2 \gamma_5 \gamma_6, \tag{2.70}$$

[10] Note that we can use the Fock basis since it is naturally the basis in which we diagonalise our Hamiltonian, therefore it is an energy basis as well.

[11] Here we denote the partial trace by $\mathrm{Tr}_{N/L}$ meaning that for the corresponding sequence of length N in the Fock Space we take the respective L elements, meaning that $N - L$ elements will be regarded as our environment.

[12] Here we have changed the original notation given in (2.4) and change it by

$$\hat{c}_{2j-1} \to \gamma_j, \qquad \hat{c}_{2j} \to \gamma_{N+j}, \tag{2.69}$$

for the case of N operators.

Note that in this way we are able to write any product of Grassmann operators. in general, for two sequences, and a fixed size L this function will be given by

$$\gamma(\vec{x},\vec{y}) = \gamma_1^{x_1}\gamma_{L+1}^{y_1}\gamma_2^{x_2}\gamma_{L+2}^{y_2}\dots\gamma_L^{x_L}\gamma_{2L}^{y_L}, \tag{2.71}$$

where this function has the property that

$$\gamma(\vec{x},\vec{y})\gamma(\vec{x}',\vec{y}') = e^{i\phi(\vec{x},\vec{y},\vec{x}',\vec{y}')}\gamma(\vec{x}+\vec{x}',\vec{y}+\vec{y}'). \tag{2.72}$$

The phase appear as a consequence of the anti-commutation relation and $\phi(\vec{x},\vec{y},\vec{x}',\vec{y}')$ is a function that would depend on the weight of the sequences $\vec{x},\vec{y},\vec{x}'$and $\vec{y}'$.
This provide a set of operators that live in the space of size L, and that will allow us to expand our operator $\hat{X}_{ij}$,[13]

$$\hat{X}_{ij} = \sum_{\vec{x},\vec{y}} f(\vec{x},\vec{y})\gamma(\vec{x},\vec{y}). \tag{2.73}$$

Our task will then be to find the coefficient $f(\vec{x},\vec{y})$, this can be achieve by multiplying on both sides by $\gamma(\vec{x'},\vec{y'})$ and taking the trace over L

$$\mathrm{Tr}_L\left(\hat{X}_{ij}\gamma^\dagger(\vec{x}',\vec{y}')\right) = \sum_{\vec{x},\vec{y}} f(\vec{x},\vec{y})\,\mathrm{Tr}_L\left(\gamma(\vec{x},\vec{y})\gamma^\dagger(\vec{x}',\vec{y}')\right). \tag{2.74}$$

The right hand side of this equation can be combined with equation (2.72) and deduce that it will give us a delta ($\delta_{\vec{x}+\vec{x}'}\delta_{\vec{y}+\vec{y}'}$). Thus the coefficients $f(\vec{x},\vec{y})$ are given by

$$f(\vec{x}',\vec{y}') = \frac{1}{2^L}\,\mathrm{Tr_L}\left(\hat{X}_{ij}\gamma^\dagger(\vec{x}',\vec{y}')\right) = \frac{1}{2^L}\left\langle\vec{n}_j\right|\gamma^\dagger(\vec{x}',\vec{y}')\left|\vec{n}_i\right\rangle. \tag{2.75}$$

Thus we first have to know how the operators γ acts over the states $|\vec{n}_i\rangle$, for this we recall the fact that the Hamiltonians are diagonalised via an orthogonal transformation which links what we call the spacial modes and the normal modes

$$\overbrace{\gamma_{i_1}\gamma_{i_2}\dots\gamma_{i_k}}^{\text{Spacial modes}} = O_{i_1\alpha_1}O_{i_2\alpha_2}\dots O_{i_k\alpha_k}\underbrace{\gamma_{\alpha_1}\gamma_{\alpha_2}\dots\gamma_{\alpha_k}}_{\text{Normal modes}}. \tag{2.76}$$

Note that this operators are diagonal over our $\vec{x}$'s and$\vec{y}$'s, which will allow us to operate over the states $|\vec{n}_i\rangle$.[14]. The equation (2.75) will be simplified to

$$\left\langle\vec{n}_j\right|\gamma(\vec{x},\vec{y})\left|\vec{n}_i\right\rangle = \delta_{\vec{n}_i+\vec{x}+\vec{y},\vec{n}_j}e^{i\phi(\vec{n}_i,\vec{n}_j,\vec{x},\vec{y})}, \tag{2.78}$$

[13]For illustrate these ideas we first suppose we have the operators in order. However, we will generalise it later.

[14]By using the function $\gamma(\vec{x},\vec{y})$ there are two ways of getting an specific state $|\vec{n}_i\rangle$

$$|n_i\rangle = \gamma(\vec{n}_i,0)\left|0\right\rangle, \qquad |n_i\rangle = \gamma(0,\vec{n}_i)\left|0\right\rangle e^{i\phi(\vec{n_i})}, \tag{2.77}$$

therefore we can say that the $\vec{x}$'s takes the 0 and turn them into a 1, and the $\vec{y}$'s take 0 and transform it into a one multiplied by a phase.

whereas the coefficients $f(\vec{x},\vec{y})$ will be given by

$$f(\vec{x},\vec{y}) = \frac{1}{2^L}\sum_{\vec{x}',\vec{y}'} \mathcal{U}_{\vec{x}\vec{x}'}\mathcal{V}_{\vec{y}\vec{y}'} \underbrace{\langle \vec{n}_j|\, \gamma(\vec{x},\vec{y})\, |\vec{n}_i\rangle}_{\propto \delta_{\vec{n}_i+\vec{n}_j,\vec{x}+\vec{y}}}. \tag{2.79}$$

In equation (2.78) the term with the delta can be change by $\delta_{\vec{n}_i+\vec{n_j},\vec{x}+\vec{y}}$, an the reason to do so is because we are working with arithmetic mod 2 and we can see that the term $\vec{n}_i+\vec{n}_j$ is nothing but the vector of differences, which means that the only values different than zero in this vector are when $n_{i_k} \neq n_{j_k}$. This result is extremely important because it tell us that when ever we work with states like $\hat{X}_{ij}$, if the vector of differences have more ones than the vector of differences $\vec{x}+\vec{y}$. Note that the vector of difference given by $\vec{x}+\vec{y}$ can have at most L errors, which means that whenever the number of ones in the difference vector defined by the states $\vec{n}_i+\vec{n}_j$ exceeds L the state $\hat{X}_{ij}$ will be immediately zero. This turns out to be an astonishing result and bring even more questions, like how likely is to have more than L errors when $N \gg L$?, what happen when we have less errors is this quantity still small enough as we expected?. These questions will be tackled in a moment but something to stress is the fact that it is the branch point of our study. We will be first addressing the first question and afterwards we will talk about the second one.

2.7.2 Codes of minimum distance constrained to an energy value

Lets recapitulate a little more what we have done in the latter chapter. We define two binary sequences of excitations $\vec{n}_i$, $\vec{n}_j$, and the vector of differences $\vec{e}_{ij} = \vec{n}_i+\vec{n}_j$, we will denote the distance between these two binary sequences by

$$\begin{aligned} d &= W(\vec{e}_{ij}) \rightarrow \text{Weight of } \vec{e}_{ij} \\ &= d_H(\vec{n}_i,\vec{n}_j) \rightarrow \text{Hamming distance.} \end{aligned} \tag{2.80}$$

we show that there is a specific distance $d > L$ at which the state $\hat{X}_{ij} = \mathrm{Tr}_{N/L}\left(\langle \vec{n}_i|\, |\vec{n}_j\rangle\right)$ is equal to zero $\hat{X}_{ij}$. This might sound quite familiar since it is connected to minimum distance codes and to see this more clearly consider a set of binary codewords

$$\mathcal{C} = \{\vec{x}^{(1)},\vec{x}^{(2)},\ldots,\vec{x}^{(2^k)}\} \quad \vec{x} \in \{0,1\}^N, \tag{2.81}$$

and size $|\mathcal{C}| = 2^{NR}$, with $R = k/N$ as defined in (2.33). Analogously, consider the Hilbert space

$$\mathcal{H}_{\mathcal{C}} = \mathrm{Span}\left(|\vec{x}^{(1)}\rangle, |\vec{x}^{(2)}\rangle, \ldots, |\vec{x}^{(2^k)}\rangle\right). \tag{2.82}$$

It is easy to check that $|\mathcal{H}_{\mathcal{C}}| = |\mathcal{C}| = 2^{NR}$. From the results showed from section 2.6 we can assure that there exist a Hilbert space with dimension $\dim|\mathcal{H}_{\mathcal{C}}| = 2^{N(1-\mathcal{H}(\ell))}$, where $\ell = L/N$, the relative distance. More specifically, $\forall\, |psi\rangle \in \mathcal{H}_{\mathcal{C}}$,

$$|\psi\rangle = \sum_{\vec{n}\in\mathcal{C}} \psi(\vec{n})\, |\vec{n}\rangle, \tag{2.83}$$

where the code $\mathcal{C}$ refers to a code (N, k, ℓ), $d > \ell$. So

$$\rho_L(\psi) = \mathrm{Tr}_{N/L}\left(|\psi\rangle\langle\psi|\right) = \sum |\psi(\vec{n})|^2 |\vec{n}\rangle\langle\vec{n}|. \tag{2.84}$$

Thus, the states belonging to $\mathcal{H}_\mathcal{C}$ are like frozen states, meaning that

$$\frac{d\rho_L}{dt} = 0. \tag{2.85}$$

One of the most important question one can ask is what is the biggest code with relative distance ℓ constrained with certain value of energy?. So if we denote by

$$S(\ell) = \bigcup_{\mathcal{C},\delta>\ell} \mathcal{H}_\mathcal{C}, \tag{2.86}$$

we would like to know how big indeed is this set and even more we would like to know its error exponent. For answering these questions we first define this problems in terms of random variables. Let $X_i(\theta_k)$ be a random probability that takes the value 1 with probability $p(\theta_k)$ and the value 0 with probability $1 - p(\theta_k)$, and let $X = \sum_i X_i(\theta_k)$ be the sum of these random variables, or equivalently the number of errors. Then we ask ourselves about the probability of having certain number of errors. Particularly we ask the probability of having a quantity of error greater of equal than a certain quantity. To address this question we can make use of the Chernoff inequality[15]

$$\begin{aligned} P(e^{SX} \geq e^{Sd}) &\leq \min_S \left\langle e^{SX}\right\rangle e^{-Sd} \\ P(e^{-SX} \geq e^{-Sd}) &\leq \min_S \left\langle e^{-SX}\right\rangle e^{Sd}. \end{aligned} \tag{2.87}$$

We then compute the expected value $\left\langle e^{SX}\right\rangle$ taking into account that we are working with independent variables

$$\left\langle e^{SX}\right\rangle = \prod_k E(e^{SX(\theta_k)}) = \prod_k \left(1 + p(\theta_k)(e^S - 1)\right) = e^{\sum_k \log(1+p(\theta_k)(e^S-1))} \equiv e^{-Nr(\delta)}, \tag{2.88}$$

where $r(\delta)$ corresponds to the correspondent error exponent

$$r(\delta) = \min_S \frac{1}{N}\left(\log\langle e^{SX}\rangle - S\delta\right). \tag{2.89}$$

Since we are interested in the case when $N \to \infty$, the error exponent can be written as

$$r(\delta) \overset{N\to\infty}{\equiv} \min_S \oint \frac{d\theta}{2\pi} \log\left(1 + p(\theta)(e^S - 1)\right) - S\delta, \tag{2.90}$$

[15] In this part we do not specify what distribution is the one we are choosing, one might guess that it is related to the Fermi-Dirac distribution, but our result will be in terms of the this general distribution $p(\theta_k)$. In the next chapter we will provide an expression for the case of the XY model.

which even though can not be analytically solved, it can be numerically solved by deriving and obtaining

$$\delta = \oint \frac{d\theta}{2\pi} \frac{p(\theta)e^S}{1-p(\theta)+p(\theta)e^S}. \tag{2.91}$$

Thus we conclude that the mean number of codes at distance d is given by

$$\langle S_{\mathcal{C}}(d)\rangle = \binom{M}{2} 2^{-Nr(\delta)}, \tag{2.92}$$

with $M \equiv 2^{NR}$, we conclude that when ever we work with rates lower than $r(\delta)/2$ the average number of pairs at a distance d goes to zero exponentially. Other wise the number of pairs that have minimum distance d are exponentially large

$$\langle S_{\mathcal{C}}(d)\rangle \doteq 2^{N(2R-r(\delta))}. \tag{2.93}$$

With the latter equation we answer the two questions asked at the end of the last section. However we have not seen what happens in the case when we work with distances less than L. In the next section we will discuss a little bit more about it and we will provide the intuition behind this case.

2.7.3 Case when the numbers of errors is less than L

For this case we know that the term $\hat{X}_{ij}$ was different than zero, however, we are going to show that nevertheless it is not necessarily zero, this quantity will be small. To study this we will consider the norm of $||\hat{X}_{ij}||$, if we take into account the relation found in (2.79) we will find that

$$\begin{aligned} ||\hat{X}_{ij}||^2 &= \mathrm{Tr}\left(\hat{X}_{ij}\hat{X}_{ij}^\dagger\right) = \frac{1}{2^L}\sum_{\vec{x},\vec{y}} |f(\vec{x},\vec{y})|^2 \\ &= \frac{1}{2^L}\sum_{\substack{\vec{x},\vec{y}\\ \vec{x}'',\vec{y}''}} \left(\mathcal{U}_{\vec{x}\vec{x}'}\mathcal{U}_{\vec{x}\vec{x}''}\right)\left(\mathcal{V}_{\vec{y}\vec{y}'}\mathcal{V}_{\vec{y}\vec{y}''}\right) e^{\phi(\vec{x},\vec{y}')\phi(\vec{x}'',\vec{y}'')}\delta_{\vec{n}_i+\vec{n}_j,\vec{x}'+\vec{y}'}\delta_{\vec{n}_i+\vec{n}_j,\vec{x}''+\vec{y}''}. \end{aligned} \tag{2.94}$$

Our purpose will then be to bound the terms $\sum_{\vec{x}}\left(\mathcal{U}_{\vec{x}\vec{x}'}\mathcal{U}_{\vec{x}\vec{x}''}\right)$, the reason to this is because if we can bound these terms by some quantity, this bound will also holds for the part containing the $\vec{y}$'s, thus for the rest of this work we will working with this quantity instead of working with (2.94). First, notice that the expression in (2.79) will not be too useful to us, since we are more interested in changing basis, for example we will be constantly changing from the spacial modes to the normal modes, so, by writing in this way the products of the operators we are not taking into account the geometric meaning of these terms. In the expansion of an operator, the resulting terms from $\gamma(\vec{x},\vec{y})$ have to be interpreted as $p-$forms, meaning that this will correspond to a volume generated by some vectors. So if we want to take into account the fact that volumes change over basis, we should have to understand that

the products of the operators in $\gamma(\vec{x},\vec{y})$ should transform in a particular way to take this into account. It is not trivial, but is simple to check that the way this products should change in order to transform as Grassmann variables and take into account the change on the volumes over transformation is via the antisymmetrizing operator. So when we see the products such as $\gamma_1\gamma_2\gamma_3$ we have to understand it as $\frac{1}{3!}\gamma_{[1}\gamma_2\gamma_{3]}$. Generalising these ideas we consider a general $p-$form, lets say $\gamma(\vec{\alpha})$

$$\gamma(\vec{\alpha}) =_{[} \gamma_{\alpha_1}\gamma_{\alpha_2}\gamma_{\alpha_2}\dots\gamma_{\alpha_p]}, \tag{2.95}$$

if each of these elements transform as $\gamma_{\alpha_i} = O_{\alpha_i j}\tilde{\gamma}_j$. The $p-$ form will transform as

$$\gamma(\vec{\alpha}) = \det\left(O\big|_{\vec{\alpha},\vec{\beta}}\right)\gamma_{\vec{\beta}}, \tag{2.96}$$

where the term $\det\left(O\big|_{alp\vec{h}a,\vec{\beta}}\right)$ refers to the minor of the matrix O. Turning back to our main problem look that the term of $\sum_{\vec{x}}(\mathcal{U}_{\vec{x}\vec{x}'}\mathcal{U}_{\vec{x}\vec{x}''})$ can be written in terms of this determinants as

$$\sum_{\vec{x}}(\mathcal{U}_{\vec{x}\vec{x}'}\mathcal{U}_{\vec{x}\vec{x}''}) = \det\left[\left(O\Pi O^T\right)\big|_{\vec{x}',\vec{x}''}\right]. \tag{2.97}$$

Therefore if we are able to find a bound for this determinant we could show that the quantity $||\hat{X}_{ij}||$ is indeed small as we have been saying. Nonetheless, to bound this quantity in general is not an easy task and we can not show a general bound for this quantity, even though, we will explore how this quantity grows for the case of the one dimensional XY model. As we will show the quantity $||\hat{X}_{ij}||$ for the case of this model is indeed small, and then it provide us an intuition that there must be some mechanism over the other models such that this quantity remains small for other Fermionic systems.
In the next chapter we are going to present the one dimensional XY model, its importance for our study as well as some calculations of the previous results to illustrate the behaviour of Ultra orthogonality for this specific case.

Chapter 3

Getting to understand Ultra Orthogonality in the XY Model.

As discussed in the previous chapters, we are interested in solvable Fermionic systems. Indeed the one dimensional XY model is one of those systems [52]. The XY Hamiltonian model is a set of N spin 1/2 particles located on the sites of d-dimensional lattice. Nevertheless, whenever we talk about the XY model, we will be referring to the $1D$ XY model.

3.1 The XY Model

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

$$H_{XY} = -\frac{1}{2}\sum_{l=0}^{N-1}\left(\frac{1+\gamma}{2}\sigma_l^x\sigma_{l+1}^x + \frac{1-\gamma}{2}\sigma_l^y\sigma_{l+1}^y + \lambda\sigma_l^z\right), \tag{3.1}$$

where γ is so-called the anisotropy parameter and represents the difference between the strength of the XX interaction and the YY interaction[1], λ is the intensity of the external magnetic field and σ_l^i is the Pauli matrix ($i = x, y, z$) acting over the l site of the chain.
The XY model is a model that has been widely studied for a variety of values of λ and γ and in some limits it has a correspondence to other models of interest in condensed matter [53–55][2].

3.1.1 The spectrum

To find the spectrum of the Hamiltonian (3.1) we need to perform some special transformations.

[1] When we talk about interactions we mean interactions between spins.

[2] Some examples of this kind, are the Boson Hubbard model in the limit of hard Bosons. the case when $\gamma = 1$ correspond to the Ising model, and the Kitaev chain is equivalent to the XY model under a proper identification of the parameters μ, t and Δ with γ and λ [53, 54]

3.1.2 Jordan-Wigner Transformation

We first consider the non local transformation given by

$$\hat{b}_l = \left(\prod_{m<l} \sigma_m^z\right) \sigma_l^-, \quad \sigma_l^- = \frac{\sigma_l^x - i\sigma_l^y}{2}, \tag{3.2}$$

where these b_l represent spinless Fermionic operators, and its canonical anticommutation relation (CAR) is given by [56]

$$\left\{\hat{b}_i^\dagger, \hat{b}_j^\dagger\right\} = \left\{\hat{b}_i, \hat{b}_j\right\} = 0, \quad \left\{\hat{b}_i^\dagger, \hat{b}_j\right\} = \delta_{i,j}. \tag{3.3}$$

So inverting the transformation we get

$$\begin{aligned} \sigma_l^z &= 1 - 2\hat{b}_l^\dagger \hat{b}_l \\ \sigma_l^x &= \left(\prod_{m<l} \left(1 - 2\hat{b}_m^\dagger \hat{b}_m\right)\right)\left(\hat{b}_l^\dagger + \hat{b}_l\right) \\ \sigma_l^y &= i\left(\prod_{m<l} \left(1 - 2\hat{b}_m^\dagger \hat{b}_m\right)\right)\left(\hat{b}_l^\dagger - \hat{b}_l\right). \end{aligned} \tag{3.4}$$

The terms of interaction in the Hamiltonian will look as

$$\begin{aligned} \hat{\sigma}_l^x \hat{\sigma}_{l+1}^x &= \left(\hat{b}_l^\dagger - \hat{b}_l\right)\left(\hat{b}_{l+1}^\dagger + \hat{b}_{l+1}\right) \\ \hat{\sigma}_l^y \hat{\sigma}_{l+1}^y &= -\left(\hat{b}_l^\dagger + \hat{b}_l\right)\left(\hat{b}_{l+1}^\dagger - \hat{b}_{l+1}\right), \end{aligned} \tag{3.5}$$

and the Hamiltonian will look like,

$$H_{XY} = -\frac{1}{2}\sum_l \left[\left(\hat{b}_{l+1}^\dagger \hat{b}_l + \hat{b}_l^\dagger \hat{b}_{l+1}\right) + \gamma\left(\hat{b}_l^\dagger \hat{b}_{l+1}^\dagger - \hat{b}_l \hat{b}_{l+1}\right)\right] - \frac{\lambda}{2}\sum_l \left(1 - 2\hat{b}_l^\dagger \hat{b}_l\right), \tag{3.6}$$

after this transformation. The term of $-\lambda N/2$ is usually ignored since it cause only a gauge of the spectrum in the energy [56] .

3.1.3 Fourier Transformation

It is possible to exploit an other symmetry in the system. It comes by considering periodic boundary conditions (PBC) [56] . This can be easily done by identifying the spin in the site N with the spin in the site 1. After imposing this condition, we have that the Fourier transform of the operator $\hat{b}_l$ will look as

$$\hat{d}_k = \frac{1}{\sqrt{N}}\sum_{l=1}^{N} \hat{b}_l e^{-i\phi_k l}, \tag{3.7}$$

with $\theta_k = \frac{2\pi}{N}k$.
Since the Fourier transformation is unitary, the operators $\hat{d}_k$ will preserve the CAR.

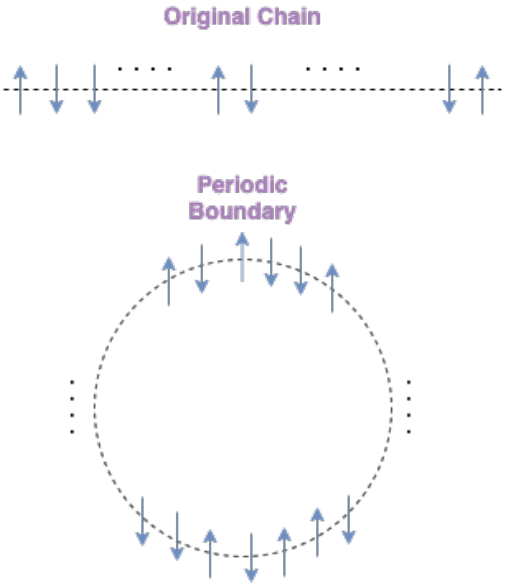


Figure 3.1: Illustration of what a boundary condition means in the case of our spin chain.

Thus, the Hamiltonian can be written in terms of the operators $\hat{d}_k$ as

$$H_{XY} = \sum_{k=-(N-1)/2}^{(N-1)/2} (-\lambda + \cos\phi_k)\, \hat{d}_k^\dagger \hat{d}_k + \frac{i\gamma}{2} \sum_{k=-(N-1)/2}^{(N-1)/2} \sin\phi_k \left(\hat{d}_k \hat{d}_{-k} + h.c\right), \tag{3.8}$$

where we have ignored an additional term which is proportional to $1/N$ which will vanish for in the thermodynamic limit $N \to \infty$ [54, 55], which is our case of interest.

3.1.4 Bogoliubov -Valantin Transformation

As mentioned before Fermionic quadratic Hamiltonians can be easily diagonalised via a Bogoliubov-Valantin transformation over the operators $\hat{d}_k$

$$\tilde{d}_k = u_k \hat{d}_k^\dagger + i v_k \hat{d}_{-k}. \tag{3.9}$$

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

$$\cos\frac{\psi_k}{2} = \frac{-\lambda + \cos\phi_k}{\sqrt{(\lambda - \cos\phi_k)^2 + (\gamma \sin\phi_k)^2}}, \tag{3.10}$$

So finally our Hamiltonian will look as

$$H_{XY} = \sum_{-(N-1)/2}^{(N-1)/2} \tilde{\Lambda}_k \tilde{d}_k^\dagger \tilde{d}_k, \tag{3.11}$$

with

$$\tilde{\Lambda}_k := \sqrt{(\lambda - \cos\phi_k)^2 + (\gamma \sin\phi_k)^2}, \tag{3.12}$$

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

3.1.5 Fermionic Covariance Matrix for the XY model

As we mentioned before, sometimes it turns out to be better, and useful to work directly with the Covariance matrix. To be able to do so, we need to express the Hamiltonian (3.1) in terms of Majoranana fermions. This can be done by using an analogous of the Jordan Wigner transformation use to diagonalised the XY Hamiltonian but now we apply it to the $2N$ Majorana fermions

$$\hat{\gamma}_l = \left(\prod_{m<l} \hat{\sigma}_m^z\right) \hat{\sigma}_l^x, \quad \hat{\gamma}_{l+N} = \left(\prod_{m<l} \hat{\sigma}_m^z\right) \hat{\sigma}_l^y, \tag{3.13}$$

where again $l = 1, 2 \ldots N-1$,

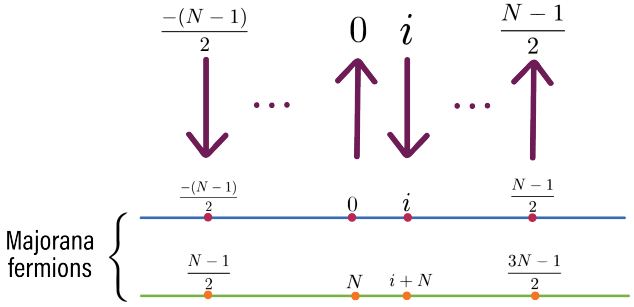


Figure 3.2: Illustration of how the spins in the chain are mapped to the Majorana fermions.

and similarly as before we have that

$$\hat{\gamma}_l \hat{\gamma}_{l+N} = \left(\prod_{m<l} \hat{\sigma}_m^z\right) \left(\prod_{m<l} \hat{\sigma}_m^z\right) \hat{\sigma}_l^x \hat{\sigma}_l^y = i\hat{\sigma}_l^z, \tag{3.14}$$

$$\hat{\gamma}_{l+N} \hat{\gamma}_{l+1} = \left(\prod_{m<l} \hat{\sigma}_m^z\right) \hat{\sigma}_l^y \left(\prod_{m<l+1} \hat{\sigma}_m^z\right) \hat{\sigma}_{l+1}^x = \hat{\sigma}_l^y \hat{\sigma}_l^z \hat{\sigma}_{l+1}^x = i\hat{\sigma}_l^x \hat{\sigma}_{l+1}^x, \tag{3.15}$$

$$\hat{\gamma}_l \hat{\gamma}_{l+N+1} = \left(\prod_{m<l} \hat{\sigma}_m^z\right) \hat{\sigma}_l^x \left(\prod_{m<l+1} \hat{\sigma}_m^z\right) \hat{\sigma}_{l+1}^y = \hat{\sigma}_l^x \hat{\sigma}_l^z \hat{\sigma}_{l+1}^y = -i\hat{\sigma}_l^y \hat{\sigma}_{l+1}^y. \tag{3.16}$$

Which coincide, up to constant factors, with the three terms in the Hamiltonian (3.1). This will lead us to a Hamiltonian of the form [31, 57]

$$H_{XY} = \frac{i}{4} \sum_{\alpha,\beta=0}^{2N} \Omega_{\alpha\beta} \left[\hat{\gamma}_\alpha, \hat{\gamma}_\beta\right], \tag{3.17}$$

where Ω is the antisymmetric matrix of the form

$$\Omega = \left[\begin{array}{c|c} 0 & \tilde{\Omega} \\ \hline -\tilde{\Omega}^T & 0 \end{array}\right], \tag{3.18}$$

with

$$\tilde{\Omega} = \begin{pmatrix} \lambda & \frac{1-\gamma}{2} & 0 & 0 & \dots & 0 & \frac{1+\gamma}{2} \\ \frac{1+\gamma}{2} & \lambda & \frac{1-\gamma}{2} & 0 & \dots & 0 & 0 \\ 0 & \frac{1+\gamma}{2} & \lambda & \frac{1-\gamma}{2} & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \dots & \vdots & \vdots \\ \frac{1-\gamma}{2} & 0 & 0 & 0 & \dots & \frac{1+\gamma}{2} & \lambda. \end{pmatrix}. \tag{3.19}$$

In general (3.18) can be diagonalised via an orthogonal transformation O [31, 57] [3]

$$\Omega = O \left[\begin{array}{c|c} 0 & \omega \\ \hline \omega & 0 \end{array}\right] O^T, \tag{3.20}$$

where $O \in O(2N)$. Writing it in terms of two smaller orthogonal matrices will look as

$$O = \left[\begin{array}{c|c} O_1 & 0 \\ \hline 0 & O_2 \end{array}\right], \tag{3.21}$$

and ω is a diagonal matrix of size $N \times N$ which holds excitation numbers $-1/2 + n$. By doing the product of matrices in (3.20) we can easily see that

$$\tilde{\Omega} = O_1 \omega O_2^T, \tag{3.22}$$

which is nothing but the singular value decomposition of the matrix $\tilde{\Omega}$. The latter result tell us that a fast way to construct the matrix O, which diagonalise Ω, is to focus on $\tilde{\Omega}$.
A fact that we can exploit is that , the matrix described in equation (3.19) $\tilde{\Omega}$ is a circulant real matrix, meaning that it can be easily diagonalised by means of a Fourier transform. So we can write

$$\tilde{\Omega}_{mn} = \frac{1}{N} \sum_{\theta_k \in (-\pi, \pi)} \omega(\theta_k) e^{\phi(\theta_k)} e^{i(m-n)\theta_k}. \tag{3.23}$$

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

$$\omega^2(\theta_k) := (\lambda - \cos\theta_k)^2 + \gamma^2 \sin^2\theta_k, \tag{3.24}$$

and

$$\phi(\theta_k) := \arctan\left(\frac{\lambda - \cos\theta_k}{-\gamma \sin\theta_k}\right). \tag{3.25}$$

[3] This special relation provide us a way to transform from spacial modes to excitation in the chain, so that we can either excite the chain and see what the spatial modes are or the other way.

So expanding the equation (3.23), we get

$$\begin{aligned}\tilde{\Omega}_{mn} &= \frac{1}{N}\left[\omega(0)+(-1)^{m-n}\omega(\pi)+2\sum_{0<\theta_k<\pi}\omega\left(\theta_k\right)\cos\left(\theta_k(m-n)+\phi\left(\theta_k\right)\right)\right] \\ &= \frac{\omega(0)}{N}+(-1)^{m-n}\frac{\omega(\pi)}{N}+\sum_{0<\theta_k\leq\pi}\omega\left(\theta_k\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right)+u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right),\end{aligned} \tag{3.26}$$

where

$$u_m^c\left(\theta_k\right)=\sqrt{\frac{2}{N}}\cos\left(m\theta_k+\phi\left(\theta_k\right)\right), \quad u_m^s\left(\theta_k\right)=\sqrt{\frac{2}{N}}\sin\left(m\theta_k+\phi\left(\theta_k\right)\right), \tag{3.27}$$

$$v_n^c\left(\theta_k\right)=\sqrt{\frac{2}{N}}\cos\left(n\theta_k\right), \quad u_n^s\left(\theta_k\right)=\sqrt{\frac{2}{N}}\sin\left(n\theta_k\right). \tag{3.28}$$

Now defining $u^s(0)=v^s(\pi)=0 \mathrm{y} u^c(0)=v^c(\pi)=\frac{1}{\sqrt{N}}$, we have that $\tilde{\Omega}_{m,n}$ can be written as

$$\tilde{\Omega}_{mn}=\sum\omega\left(\theta_k\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right)+u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right). \tag{3.29}$$

Therefore the upper part of the Hamiltonian reads

$$H=\sum_{m,n=0}^{N-1}\frac{i}{4}\sum_{\theta_k=0}^{\pi}\omega\left(\theta_k\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right)+u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right)\left[\hat{\gamma}_n,\hat{\gamma}_{m+N}\right], \tag{3.30}$$

$$H=\sum_{\theta_k=0}^{\pi}\omega\left(\theta_k\right)\left(\underbrace{\left[\hat{\gamma}_k^c,\hat{\gamma}_{k+N}^c\right]}_{1-2\sigma_k^z}+\underbrace{\left[\hat{\gamma}_k^s,\hat{\gamma}_{k+N}^s\right]}_{1-2\sigma_k^z}\right), \tag{3.31}$$

where

$$\hat{\gamma}_k^{c,s}:=\sum_n u_n^{c,s}\left(\theta_k\right)\hat{\gamma}_n, \quad \hat{\gamma}_{k+N}^{c,s}:=\sum_n v_n^{c,s}\left(\theta_k\right)\hat{\gamma}_{n+N}. \tag{3.32}$$

Now we look back on the fact that the Fermionic covariance matrix, defined by $\Gamma_{\alpha\beta}=\frac{1}{2i}\operatorname{tr}\left(\rho\left[\gamma_\alpha,\gamma_\beta\right]\right)=\frac{1}{2i}\langle[\gamma_\alpha,\gamma_\beta]\rangle$, that brings Ω into its Williamson form, does the same on the Fermionic covariance matrix. Thus for a state $|\vec{n}\rangle$, consider an eigenstate of the base (c,s,θ_k), where $m^{c,s}\left(\theta_k\right)-1/2$, with $n^{c,s}\left(\theta_k\right)$ the occupation number of *cosine*, *sine* in the $k-$mode. We get

$$\begin{aligned}\tilde{\Gamma}_{mn} &= \sum_{\theta_k}^{\pi}\left[m^c\left(\theta_k\right)u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right)+m^s\left(\theta_k\right)u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right] \\ &= \sum_{\theta_k}^{\pi}\left(\frac{m^c\left(\theta_k\right)+m^s\left(\theta_k\right)}{2}\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right)+u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right) \\ &+ \sum_{\theta_k}^{\pi}\left(\frac{m^c\left(\theta_k\right)-m^s\left(\theta_k\right)}{2}\right)\left(u_m^c\left(\theta_k\right)v_n^c\left(\theta_k\right)-u_m^s\left(\theta_k\right)v_n^s\left(\theta_k\right)\right).\end{aligned} \tag{3.33}$$

by defining $m^{\pm}(\theta_k) = \frac{m^c(\theta_k) \pm m^s(\theta_k)}{2}$ and inverting the transformations done above, we finally get that

$$\tilde{\Gamma}_{mn} = \overbrace{\sum_{\theta_k}^{\pi} m^{+}(\theta_k) e^{i\phi(\theta_k)} e^{i(n-m)\theta_k}}^{\tilde{\Gamma}^{+}_{mn}} + \underbrace{\sum_{\theta_k}^{\pi} m^{-}(\theta_k) e^{i\phi(\theta_k)} e^{i(n+m)\theta_k}}_{\tilde{\Gamma}^{-}_{mn}}. \tag{3.34}$$

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

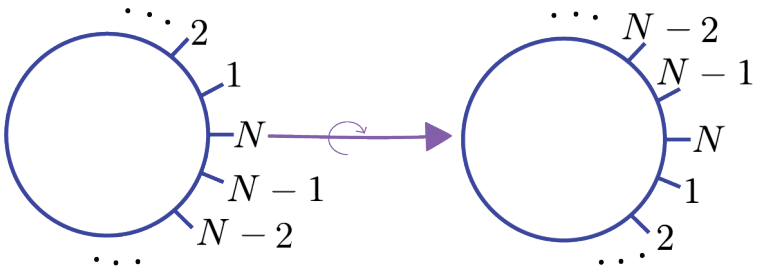


Figure 3.3: Meaning of the relabel done in the circulant matrix, which can be seen as a reflection over the circle.

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

$$\begin{pmatrix} a_0 & a_{-1} & \cdots & a_2 & a_1 \\ a_1 & a_0 & \cdots & a_3 & a_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{-1} & a_{-2} & a_{-3} & \cdots & a_0 \end{pmatrix}, \tag{3.35}$$

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

$$\begin{pmatrix} a_0 & a_1 & \cdots & a_{-2} & a_{-1} \\ a_1 & a_2 & \cdots & a_{-1} & a_0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{-1} & a_0 & a_1 & \cdots & a_{-2} \end{pmatrix}. \tag{3.36}$$

So we can spot 3 things. First, the FCM always can be written as a circulant matrix plus an anticirculant matrix. Second, in the Ground state, the FCM is circulant only, since the fermion occupation numbers $n^c(\theta_k) = n^s(\theta_k) = 0, \forall k$. Third, for a generic state, we have that in average the FCM matrix is always circulant, because $\langle n^c(\theta_k) \rangle = \langle n^s(\theta_k) \rangle$.
Now that we have showed a full characterization of the XY model, we change gears and start

to board our main problem. In the next section we are going to show how this is possible to find a bound in the case of the XY model and even more, we will show that the mechanism for Ultra Orthogonality in this system, is indeed related with the structure of the system.

3.2 Exploring "Ultra Orthogonality"

Now having all the tools we need we can apply what we discussed in section 2. First we are going to look at the error exponents for the XY model, we will compute the probabilities of having errors and show some numerical results. Then we move to our next problem, we will show for the XY model is possible to bound the determinant and even more we find an analytical result for this determinant based on a generalisation of Szegö limit theorems.

3.2.1 Error exponent

First we recall the fact that in order to sample states at a given temperature we make use of a very well known technique, named Gibbs sampling. This technique is a special case of a Markov chain Monte Carlo algorithm. Is is well known for obtaining sequences of observations which are approximated from a specified multivariate probability distribution, It is a very useful tool for simulations in Markov processes for which the transition from which transition matrix cannot be formulated explicitly because the state space is too large [58–60]. Our goal is then to sample over excited states $|\vec{n}\rangle$, where the occupations n_q will be sampled independently according to the Boltzmann distribution

$$p(\vec{n} \mid \beta) = \prod_{q=0}^{N-1} p\left(n_q \mid \beta\right), \quad p\left(n_q \mid \beta\right) = \frac{e^{-\beta\epsilon(\theta_q)n_q}}{\sum_{n_q} e^{-\beta\epsilon(\theta_q)n_q}}, \tag{3.37}$$

where we explicitly put the dependence of the energy with certain angle in (3.37) having in mind the case of the XY model[4]. The goal of using this technique that the size of the Hilbert space is to find the probability distribution with a given temperature β. It is well known that in the Fermionic case, the average number of excitations in the mode at angle θ is given by

$$f\left(\theta_q \mid \beta\right) \equiv \langle n_q \rangle_\beta = \frac{1}{e^{\beta\epsilon(\theta_q)} + 1}, \tag{3.38}$$

while the variance in the number of excitations is given by

$$v\left(\theta_q \mid \beta\right) \equiv \left\langle n_q^2 \right\rangle_\beta - \langle n_q \rangle_\beta^2 = \frac{1}{e^{\beta\epsilon(\theta_q)} + 1} = f\left(\theta_q\right)\left(1 - f\left(\theta_q\right)\right). \tag{3.39}$$

[4]However, the XY model is not the only model in which the energy can be parametrized in terms of an angle, in general it has been shown that every one-dimentional translationally invariant closed chain of free Fermions/Bosons will have this property [32, 52, 53, 57, 61].

With this method of sampling, the mean energy is

$$\langle E\rangle_\beta = \sum_q f_q\left(\theta_q \mid \beta\right)\epsilon_q\left(\theta_q\right) = N\oint_N \frac{d\theta}{2\pi} f(\theta \mid \beta)\epsilon(\theta), \tag{3.40}$$

where $\oint$ denote the Riemann sum approximation to the respective integral with N subdivisions. As we mentioned before we are interested in the limit $N \to \infty$, we will replace the sums by its correspondent integral. Similarly we can compute the energy variance of the sampled states

$$\left\langle \Delta E^2\right\rangle_\beta = \sum_q v_q\left(\theta_q \mid \beta\right)\epsilon_q^2\left(\theta_q\right) = N\oint_N \frac{d\theta}{2\pi} v(\theta \mid \beta)\epsilon^2(\theta). \tag{3.41}$$

Thus Gibbs sampling provides an even sampling of states within ΔE of the energy $\langle E\rangle$, where $\Delta E/\langle E\rangle \sim O(N^{-1/2})$.
Thinking back on the case of the XY model, the ensemble defined by Gibbs sampling is nothing but the canonical ensemble defined on the full chain, with thermal density matrix

$$\rho_T(\beta, N) = \sum_{\vec{n}} p(\vec{n} \mid \beta)|\vec{n}\rangle\langle\vec{n}| = \frac{e^{-\beta H_N}}{Z(\beta, N)}, \quad \log Z(\beta, N) = N\oint_N \frac{d\theta}{2\pi}\log\left(1 \pm e^{-\beta\epsilon(\theta)}\right). \tag{3.42}$$

This thermal state defines a reduced density matrix in the subchain of length L

$$\rho(\beta, N)|_L \equiv \mathrm{Tr}_{N-L}\,\rho_T(\beta, N), \tag{3.43}$$

which is expected to correspond to the local thermal state,

$$\rho(\beta, N)|_L \simeq \rho_T(\beta, L) \equiv \frac{e^{-\beta H_L}}{Z(\beta, L)}, \tag{3.44}$$

for L sufficiently larger in comparison to the correlation length, in which the boundary effects can be neglected. Since we know this states conserve Gaussianity under partial traces, the reduced state $\rho(\beta, N)|_L$ will be also Gaussian, meaning that the state will be uniquely characterised by its covariance matrix. This arguments lead us to the conclusion that the Gibbs average of the reduced partial density matrices $\rho_L(\vec{n})$ satisfies

$$\rho(\beta, N)|_L = \langle\rho_L(\vec{n})\rangle_\beta\,. \tag{3.45}$$

The latter series of arguments where need to understand why this method of sampling is appropriate for our purpose. Even more, it provide us an expression to the probability of having and excitation (a 1) or not (0). Having said that, we move gears to the computing the probability of having an error over two independent sequences. We can write this probability in terms of our energy as and a given θ_k

$$n(\theta_k) = \frac{1}{1 + e^{\beta(\Omega(\theta_k) - \Omega^*)}}, \tag{3.46}$$

where $\Omega(\theta_k)$ is given by equation (3.12), and Ω^* correspond to the minimum of energy[5]. Then the probability of two sequences not having an error in the position k will be given by

$$1 - p(\theta_k) = n(\theta_k)^2 + (1 - n(\theta_k))^2 = \frac{1}{1 + \operatorname{sech}\left(\beta(\Omega(\theta_k) - \Omega^*)\right)}, \tag{3.47}$$

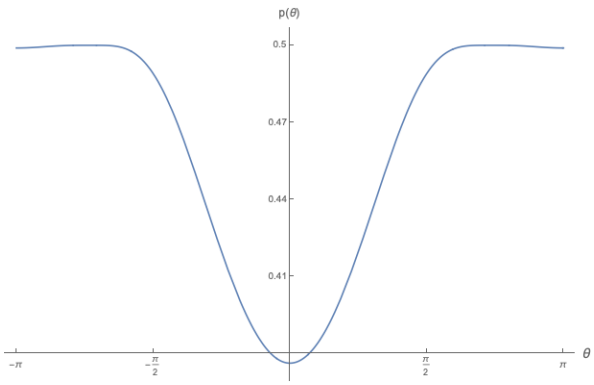
and then the probability of having an error will be given by

$$p(\theta_k) = \frac{1}{1 + \cosh\left(\beta(\Omega(\theta_k) - \Omega^*)\right)}, \tag{3.48}$$

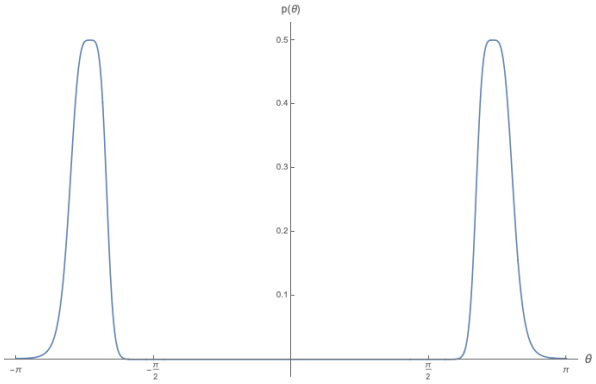
So the expected value of the number of errors $X = \sum_k X(\theta_k)$ is given by

$$\mu = \sum p(\theta_k) \overset{N\to\infty}{=} \frac{N}{2\pi} \oint d\theta \frac{1}{1 + \cosh\left(\beta(\Omega(\theta) - \Omega^*)\right)}. \tag{3.49}$$

In Figure 3.4, we show the behaviour of the probability distribution as a function of the angle.



(a) Probability distributions of errors as a function of the position in the chain. this plot correspond to a value of $\beta = 1$.



(b) Probability distributions of errors as a function of the position in the chain. this plot correspond to a value of $\beta = 80$.

Figure 3.4: An illustration of how the probability distribution $p(\theta_k)$ behaves as a function of the angle θ. As we can see in both figures there are regions of the chain in which the probability of having an error is smaller.

As we would like to study a little deeper the behaviour of the expected value as a function of the temperature we could take the obtain some expressions for some limit cases. We can expand (3.49) when $\beta \to 0$.

[5]It is not hard to check that the minimum of energy happens at $\theta^* = \pm \operatorname{acos}\left(-\frac{\lambda}{1-\gamma^2}\right)$ with value

$$\sqrt{\frac{\gamma^2(\gamma^2 - 1 + \lambda^2)}{(\gamma^2 - 1)}}.$$

$$\mu \approx \frac{N}{2\pi}\left(\frac{2\pi}{2} - \beta^2 \int_{-\pi}^{\pi} \frac{(\Omega(\theta)-\Omega^*)^2}{8} d\theta\right)$$

$$= \frac{N}{2} - \frac{N\beta^2}{16}\int_{-\pi}^{\pi}(\Omega(\theta)-\Omega^*)^2 d\theta, \tag{3.50}$$

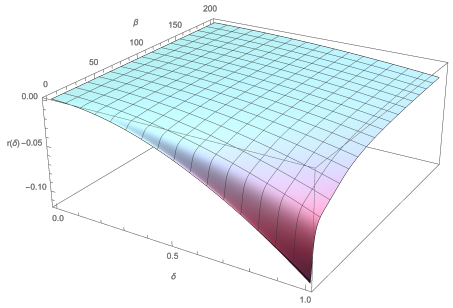
which means that in the limit of $\beta \to 0$ (High temperatures), the expected value μ will tend to $N/2$. From other side, if we look at the case when $\beta \to \infty$, the integral (3.49) can be approximated as

$$\mu \to \frac{N}{2\pi}\oint e^{-\beta(\Omega(\theta)-\Omega^*)} d\theta \to 0, \tag{3.51}$$

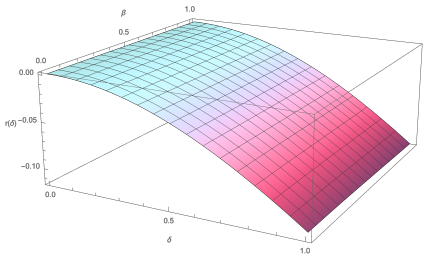
which is agreement with the fact that at zero temperature, the possible sequences are those which have no excitations, and therefore will have only zeros.
We can now compute the error exponent associated with this model by just replacing the probability in (3.48) into the expression found in (2.90), to obtain

$$r(\delta) = \min_S \oint \frac{d\theta}{2\pi} \log\left(1 + \frac{1}{1+\cosh\left(\beta(\Omega(\theta)-\Omega^*)\right)}(e^S - 1)\right) - S\delta. \tag{3.52}$$

Even though we are not able to find an exact solution to this, we can do a numerical exercise and see how this behaves as a function of the temperature. In figure 3.5 our results are shown.



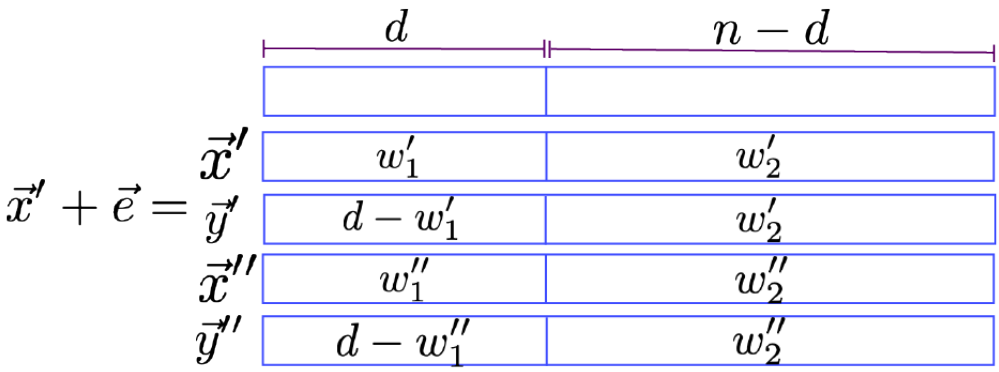
(a) Error exponent function for the XY model, in the figure the behaviour as a function of β is shown. The equation(3.52) is solved for values of β running from 0.01 to 200



(b) Zoom on the region of lower β's. The equation(3.52) is solved for values of β running from 0.01 to 1

Figure 3.5: Plot of the solution of equation (3.52), the behaviour of $r(\delta)$ is shown as a function of β.

Now we go back the problem we left behind and start analysing the case when the number of errors is less than L.



3.2.2 Finding Bounds for $\hat{X}_{ij}$

In Chapter two we discussed that if we are able to show that the operator $\hat{X}_{ij}$ in norm is small, we could assert that Super Orthogonality will hold. First we want to find a bound to this quantity in terms of the transformations $\mathcal{U}$ and $\mathcal{V}$. As a first approach we could reorder our sequences and divide them in two, one part of length d which will have all the errors and the rest $N-d$. Note that if we do so, for the region of the errors, one can check that if we have in the $\vec{x}'$ a 0 then in the sequence $\vec{y}'$ there must be a 1 in the same position, and for the region of the $N-d$ we need that whenever we have a 0 in the $\vec{x}'$ sequence, there must be a 0 in the $\vec{y}'$. Indeed it holds also for $\vec{x}''$ and $\vec{y}''$, which in general provides a relation with the weights of each pair of sequences. In figure **??** we graphically describe what we mentioned above.

Como discutimos la anterior vez, queremos probar que incluso en el caso en el que tuviéramos dos secuencias para las cuales el numero de errores fuera menor a L, el producto cruzado debe de ser de igual forma pequeño. Entonces para esto podemos estudiar

$$||X_{12}||_2^2 = \mathrm{Tr}\, X_{12} X_{12}^\dagger \tag{3.53}$$

Como lo se discutió en la ultima sesión, esta norma puede ser escrita como:

$$||X_{12}||_2^2 = \frac{1}{2^L} \sum_{\vec{x},\vec{y}} (O_{\vec{x}\vec{x}'} O_{\vec{x}\vec{x}''})(V_{\vec{y}\vec{y}'} V_{\vec{y}\vec{y}''}), \tag{3.54}$$

no obstante, veíamos que al interpretar estas secuencias $\vec{x}$ como $p-$formas, se tenía que la si se quería que esto transformara como una $p-$forma y que además los operadores transformaran como variables de Grassmann, es necesario que estas transformaciones se efectuaran por medio de un determinante, con esto teníamos entonces que

$$\underbrace{\sum_{\vec{x}}}_{\text{Suma sobre p-formas}} (O_{\vec{x}\vec{x}'} O_{\vec{x}\vec{x}''}) = \det\left[(O^T \Pi_L O)\big|_{\vec{x}',\vec{x}''} \right],$$

en donde Π_L es el proyector al espacio L de la cadena y $|_{\vec{a},\vec{b}}$ representan el menor sobre $\vec{a}$, $\vec{b}$. Como primera aproximación se efectuó un análisis de que tan grande es este determinante en términos de los grados de las $p-$formas, o como se conoce en teoría de información, el peso de la secuencia. Con esto llegábamos entonces a plantear una posible cota en términos de la cota de Haddamard para el determinante. As í se tenía entonces que

$$||X_{12}||_2^2 \leq \frac{1}{2^L} \sum_{w_1,w_2} \binom{d}{w_1}^2 \binom{n-d}{w_2}^2 \left[\frac{2}{N}(w_1+w_2)\right]^{w_1+w_2} \left[\frac{2}{N}(d-w_1+w_2)\right]^{d-w_1+w_2}, \tag{3.55}$$

con,

$$\begin{aligned} 0 \leq w_1 \leq d \\ 0 \leq w_2 \leq N-d \\ 0 \leq w_1+w_2 \leq L \\ 0 \leq d-w_1+w_2 \leq L, \end{aligned} \tag{3.56}$$

y d el numero de errores. Si suponemos que esto sigue una ley de grandes desviaciones, tenemos entonces que

$$||X_{12}||_2^2 \leq e^{Nf(w_1,w_2,L,d,N))}, \tag{3.57}$$

y

$$\begin{aligned} f(w_1,w_2,L,d,N) = 2\frac{d}{N}H_2\left(\frac{w_1}{2}\right) + 2\left(1-\frac{d}{N}\right)H_2\left(\frac{w_2}{N-d}\right) + \frac{w_1+w_2}{N}\log_2\left(2\frac{w_1+w_2}{N}\right) \\ + \frac{d-w_1+w_2}{N}\log_2\left(2\frac{d-w_1+w_2}{N}\right) - \frac{L}{N} \end{aligned} \tag{3.58}$$

Desafortunadamente como se muestra en las figuras **??** y **??** el comportamiento de f sobre la región descrita en (3.56) hace que no sea posible asegurar que existirá una cota pequeña para distintos valores de N, L, d

Bibliography

[1] Ludwig Boltzmann. *Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie: (vorgelegt in der Sitzung am 8. Februar 1866)*. Staatsdruckerei. Google-Books-ID: 8Hc5AAAAcAAJ. iv, 3

[2] 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. 1

[3] Sheldon Goldstein, Joel L. Lebowitz, Roderich Tumulka, and Nino Zanghì. Canonical typicality. 96(5):050403. 1

[4] Sandu Popescu, Anthony J. Short, and Andreas Winter. Entanglement and the foundations of statistical mechanics. 2(11):754–758. 1, 12, 14

[5] L. D. Landau and E. M. Lifshitz. *Statistical Physics: Volume 5*. Butterworth-Heinemann, 3 edition edition. 1

[6] Sandu Popescu, Anthony J. Short, and Andreas Winter. The foundations of statistical mechanics from entanglement: Individual states vs. averages. 1, 12, 14

[7] Noah Linden, Sandu Popescu, Anthony J. Short, and Andreas Winter. Quantum mechanical evolution towards thermal equilibrium. 79(6):061103. 2, 15, 16, 17, 18, 19, 20

[8] J. Uffink. *Handbook for Philsophy of Physics 924-1074*. Elsevier, Amsterdam. 2, 5

[9] Ludwig Boltzmann. Zur priorität der auffindung der beziehung zwischen dem zweiten hauptsatze der mechanischen wärmetheorie und dem principe der kleinsten wirkung. 219(6):211–230. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18712190606. 3

[10] Ludwig Eduard Boltzmann. *Einige allgemeine sätze über Wärmegleichgewicht*. K. Akad. der Wissensch. 3

[11] Josiah Willard Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. C. Scribner's Sons. 3

[12] Paul 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]. 5

[13] César R. de Oliveira and Thiago Werlang. Ergodic hypothesis in classical statistical mechanics. 29(2):189–201. Publisher: Sociedade Brasileira de Física. 5

[14] Larry Engelhardt, Marie Lopez del Puerto, and Norman Chonacky. Simple and synergistic ways to understand the boltzmann distribution function. 83(9):787–793. Publisher: American Association of Physics Teachers. 5

[15] Giovanni Gallavotti. *Statistical Mechanics: A Short Treatise*. Theoretical and Mathematical Physics. Springer-Verlag. 5

[16] Sheldon 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. 5

[17] Kerson Huang. *Statistical Mechanics, 2nd Edition*. John Wiley & Sons, 2 edition edition. 6, 7

[18] E. Schrödinger. Discussion of probability relations between separated systems. 31(4):555–563. Publisher: Cambridge University Press. 9

[19] Artur Ekert, Patrick Hayden, and Hitoshi Inamori. Basic concepts in quantum computation. version: 1. 9

[20] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press. Google-Books-ID: 65FqEKQOfP8C. 9

[21] K. W. Edwin and R. J. Taud. Optimal operating system composition using integer programming and statistic reserve determination. In *Electrical Load-Curve Coverage*, pages 13–30. Pergamon. 9

[22] John Von Neumann. *Mathematical foundations of quantum mechanics*. Princeton, N.J. : Princeton University Press. 10

[23] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 1 edition edition. 11, 32, 34, 38

[24] 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. 11, 31

[25] Martin B. Plenio and S. Virmani. An introduction to entanglement measures. 11

[26] J. Gemmer and G. Mahler. Distribution of local entropy in the hilbert space of bi-partite quantum systems: origin of jaynes' principle. 31(2):249–257. 11

[27] Don N. Page. Average entropy of a subsystem. 71(9):1291–1294. 11

[28] Paolo Facchi and Giancarlo Garnero. Quantum thermodynamics and canonical typicality. 14(8):1740001. 14

[29] 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. 18

[30] Giancarlo Camilo, Gabriel T. Landi, and Sebas Eliëns. On the strong subadditivity of the r\'enyi entropies for bosonic and fermionic gaussian states. 99(4):045155. 19

[31] Alonso Botero and Benni Reznik. BCS-like modewise entanglement of fermion gaussian states. 331(1):39 – 44. 23, 24, 51, 52

[32] Eduardo Fradkin. *Field Theories Of Condensed Matter Systems Vol#82 On-demand Printing Of #52253*. CRC Press, 1 edition edition. 24, 55

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

[34] N. N. Bogoljubov. On a new method in the theory of superconductivity. 7(6):794–805. 24

[35] B. Westwanski and A. Pawlikowski. On a general statistical wick theorem. 43(2):201–202. 24, 25

[36] Luca Guido Molinari. Notes on wick's theorem in many-body theory. 24

[37] Martin Leijnse and Karsten Flensberg. Introduction to topological superconductivity and majorana fermions. 27(12):124003. 24

[38] Siew-Ann Cheong and Christopher L. Henley. Many-body density matrices for free fermions. 25

[39] Christina V. Kraus, Michael M. Wolf, J. Ignacio Cirac, and Geza Giedke. Pairing in fermionic systems: A quantum information perspective. 79(1):012306. 26

[40] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. 409(6816):46–52. Number: 6816 Publisher: Nature Publishing Group. 27

[41] Barbara M. Terhal and David P. DiVincenzo. Classical simulation of noninteracting-fermion quantum circuits. 65(3):032325. 27

[42] E. Knill. Fermionic linear optics and matchgates. 27

[43] 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. 27

[44] D. E. Soper. Construction of the functional-integral representation for fermion green's functions. 18:12. Institution: Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403. 28

[45] Sergey Bravyi. Classical capacity of fermionic product channels. version: 1. 28

[46] Sergey Bravyi. Lagrangian representation for fermionic linear optics. 29

[47] E. T. Jaynes. *Probability Theory: The Logic of Science*. 32, 38

[48] A. Barg and G.D. Forney. Random codes: minimum distances and error exponents. 48(9):2568–2573. Conference Name: IEEE Transactions on Information Theory. 38, 39

[49] R. Gallager. Low-density parity-check codes. 39

[50] Robert G. Gallager. *Information Theory and Reliable Communication*. 39

[51] R. Gallager. The random coding bound is tight for the average code (corresp.). 19(2):244–246. 39

[52] Elliott Lieb, Theodore Schultz, and Daniel Mattis. Two soluble models of an antiferromagnetic chain. 16(3):407–466. 48, 55

[53] Shigetoshi Katsura. Statistical mechanics of the anisotropic linear heisenberg model. 127(5):1508–1518. 48, 55

[54] Eytan Barouch and Barry M. McCoy. Statistical mechanics of the XY model. II. spin-correlation functions. 3(2):786–804. 48, 50

[55] Eytan Barouch, Barry M. McCoy, and Max Dresden. Statistical mechanics of the XY model. i. 2(3):1075–1092. 48, 50

[56] A. F. Reyes-Lega. Some aspects of operator algebras in quantum physics. pages 1–74. 49

[57] J. I. Latorre, E. Rico, and G. Vidal. Ground state entanglement in quantum spin chains. 4(1):48–92. 51, 52, 55

[58] Christian P. Robert and George 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. 55

[59] Markov chain monte carlo in practice. 55

[60] Gibbs sampler - an overview | ScienceDirect topics. 55

[61] J. Eisert, M. Cramer, and M. B. Plenio. Area laws for the entanglement entropy - a review. 82(1):277–306. version: 4. 55

[62] Ingo Peschel. Calculation of reduced density matrices from correlation functions. 36(14):L205–L208.

[63] *Field States*, chapter 11, pages 291–319. John Wiley & Sons, Ltd, 2005.

[64] Philippe Francesco, Pierre Mathieu, and David Sénéchal. *Conformal Field Theory*. Graduate Texts in Contemporary Physics. Springer-Verlag.

[65] Neri Merhav. Error exponents of typical random codes.

[66] Yuval Domb, Ram Zamir, and Meir Feder. The random coding bound is tight for the average linear code or lattice. 62(1):121–130. Conference Name: IEEE Transactions on Information Theory.

[67] Ming-Chiang Chung and Ingo Peschel. Density-matrix spectra of solvable fermionic systems. 64(6):064412.

[68] Yuezhen Niu, Suk Bum Chung, Chen-Hsuan Hsu, Ipsita Mandal, S. Raghu, and Sudip Chakravarty. Majorana zero modes in a quantum ising chain with longer-ranged interactions. 85(3):035110.