university of groningen / faculty of science and engineering / mathematics and applied mathematics

Ising Model and Spin Glasses

Bachelor's Project Mathematics and Phyisics

July 2022

Student: F. Cornia

First supervisor: Prof.dr. P. Trapman

First supervisor: Prof.dr. E. Pallante

Abstract

The Ising model is a statistical model that can be used to describe the energy of a system of atoms arranged on a lattice, where the interaction between these atoms is governed by the interaction among their spin and a possible external magnetic field. This model and some more complicated version of it can then be used to described the behavior of spin glasses. These are systems where the spin of the atoms is randomly oriented in space. The study of spin glasses has brought new techniques to solve combinatorial optimization problems and we will see and example of how to do so, by looking at the Travelling Salesman Problem. We will then look at how these problems can be solved both classically and the aid of a quantum computer. We will conclude by giving a brief explanation of what quantum computers are, how they differ from the classical ones, their potential and current prediction for the future developments.

Contents

1 Introduction

The Ising model is named after the mathematician Ernst Ising that in 1924 did his PhD thesis on this model. Ising's goal was to prove that the model does not present a phase transition in the one dimensional case, this is true and he managed to show it. After his finding then Ising incorrectly generalized what he had proven to higher dimensions. It was not until 1936 when Peierls showed that indeed for the two dimensional case we have a phase transition in the presence of a small external magnetic field, he mainly looked at a square lattice but the result can be generalized. Later in 1944 Onsager proved the phase transition in two dimensions in absence of the external magnetic field. This thesis is going to focus on two main topics: the Ising model and Spin glasses. The reason why we want to look into the Ising model is that this is a statistical model that has been successfully used to describe systems of atoms arranged on a regular lattice and carrying a magnetic moment (spin) that influences the interaction between them. We will then consider different version of the Ising model and a particular one, the Random Ising Model, will be used to describe spin glass. The motivation behind the study of spin glasses is that their study has brought application in many fields of science. The first time spin glasses were studied the name spin glass was used to indicate a metallic alloy of gold with small traces of iron, these alloys had peculiar magnetic properties. The magnetic susceptibility presented a peak that seemed to indicate a phase transition. By looking at the specific heat of the material it should have shown a peak as well around the same area of that of the magnetic susceptibility. This was not the case and it implied that a peculiar type of phase transition occurred in these sorts of systems. The knowledge that derived from this was then applied to many areas of science and among these there is also the one we will mainly treat in this thesis. Our specific reason to study spin glasses is their relation to combinatorial optimization problems and how models, as the Ising model used to describe them, can be applied to find new ways to solve optimization problems. Our goal in this thesis is to give a description of the Ising model and some of its variation and show how these can be applied to spin glasses first and optimization problem then. These connections are well known already and are known to be of great importance in the area of complexity theory, a field for which half of the Nobel prize of 2021 was awarded to Giorgio Parisi. We then want to show how optimization problems can be solved with the help of quantum computers.

To do what we just described we will make use of basic concepts of probability theory such as the notion of what a random variable and a measure on a probability space are this latter definition is not key to the whole thesis, it is mainly used in chapter 5.5 when we describe the Random Cluster Model. Some other knowledge required is the concept of Hamiltonian, that in this thesis is used as being the function that describes the energy of a system. Linked to the concept of Hamiltonian we have that of the Partition function. The partition

function Z can be described using the formula

$$Z = \sum_j \exp\{-\beta E_j\}$$

where $\beta = \frac{1}{kT}$ is the inverse temperature. This partition function is by definition the summation over all possible states of the system of the exponential of $-\beta$ times the energy of the state.

Outline

The thesis will start by introducing the reader to the Ising model and this is done in chapter 2. Here we treat mainly the one dimensional case and give the basics to be able to tackle the rest of the thesis. Chapter 3 then treats the higher dimensional case and gives an idea on how we can show the existence of a phase transition for dimensions 2 or higher, by use of the mean field approximation. The next chapter 4 then shows the formal proof of a phase transition for the case of the two dimensional square lattice, and we then conclude the first part of the thesis in chapter 5 by introducing the reader to an example of an infinite dimensional lattice, the Bethe lattice, and a more complicated version of the Ising model, the Random Cluster Model. The second part starts in chapter 6 where we introduce spin glasses and we describe their basic physical properties and show their link to the Ising model. In this chapter we also introduce a variation of the Ising model, the Random Ising model that will be used later as well. Chapter 7 shows the link between all the previously mentioned topics and combinatorial optimization problem; and then gives an example on how we can do the mapping from one to the other. In the first part of this chapter we also introduce the reader to the quite wide topic of P versus NP problems. We then conclude the thesis with the last two chapter 8 and 9, where we first describe a couple of algorithms that can be used to solve optimization problems, ad we then describe how quantum computers can be used to implement new types of algorithms that might be able to solve problems, too hard for classical ones. We conclude the thesis with a brief outline of the prediction of what we might expect for the next decade in the area of quantum computers.

Acknowledgements

Before I start digging into the topic I would like to take a moment to thank both my supervisors Prof.dr. E. Pallante and Prof.dr. P. Trapman. I really appreciate them guiding me into this topic and giving me the chance to pursue this as my research topic. I also would like to thank them for always being available for my questions and the insightful answers they gave me. To conclude I would like to thank the University of Groningen and the faculty of science and engineering for allowing me to use the spaces they had reserved on campus.

2 One-dimensional Ising Model

We can start by describing the easiest example of the Ising model, which is the one dimensional case. In this model we start by taking a series of discrete points on a line, equally spaced dots for example.

At each dot we can assign one of two possible values either 1 or -1. This is sometimes represented by spin 'up' or 'down', usually visualized by arrows, see figure 1.

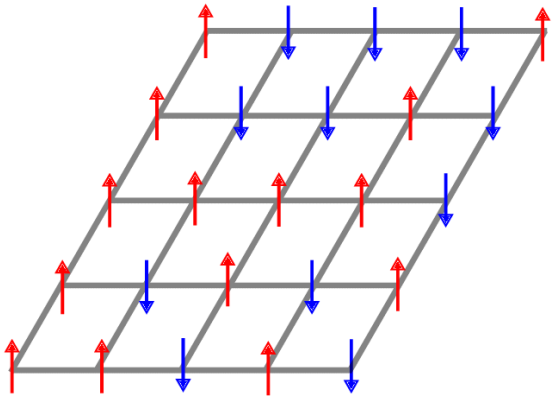

Figure 1: The figure represents a set of atoms placed on the lattice sites, arranged in a two dimensional square lattice, the spin of each atom can either be in the 'up' state, in this case represented by the red colour, or the 'down' state, in blue [Wald, 2017]

Once we set up this part of the model we now have that each point interacts with its nearest neighbours, two in the one dimensional case. In this thesis we will assume, unless specified otherwise, that the spin of the points tend to align, which means that the values assigned to each dot tend to be the same; this one is the so called ferromagnetic case, opposed to the antiferromagnetic case where the spins tend to anti-align.

If we imagine the dots we have as atoms in a material and the value assigned to each of them as the magnetic moment of this atom, we can start to define the energy of the system we are describing.

In describing the energy we start by the easier case where we place an external magnetic field B, ignoring for the moments the interaction among the atoms, we can then describe the energy each point brings to the whole system as

$$E_i = \sigma_i \mu h \tag{1}$$

where we are taking h to be just the parallel component of the magnetic field in the direction of the spin or magnetic moment of the atoms. The number μ is just a constant, representing the strength of the magnetic moment, and σ_i represent the spin of each atom, so it has a value of ± 1.

To be able to obtain a function describing the energy of our system we first need to find the partition function.

The partition function assumes different meaning based on the areas of science under which it is studied. What is meant by this is that in statistics the partition function has the role of a normalizing constant, it is needed to normalize the probability distribution.

In physics it is expressed numerically in the same way but it represents, not a sole constant but it is linked to properties of the system, such as Entropy and energy.

In both cases the partition function is expressed as follows

$$Z = \sum_j \exp\{-\beta E_j\} \tag{2}$$

A few terms need to be explained. β is for this moment assumed to be constant and follows the relation $\beta = \frac{1}{kT}$ where T is the temperature of the system and k is the Boltzman constant. The sum is over all possible configurations j of the system, hence E_j is the energy of the specific configuration j.

Back to the one-dimensional Ising model, given a number N of points the total number of possible configurations is 2^N, and the number of configuration with a fixed number n of 'up' spins and n 'down' spins is $\frac{N!}{n!\cdot m!}$. And if we look at the energy of such a system with n 'ups' and m 'downs' given the energy for each particle as in equation (1), to get the total energy we sum all the individual ones and get

$$E = \sum_i \sigma_i \mu h = (n - m)\mu h \tag{3}$$

where the sum is over all lattice sites i.

With this result we can now find the partition function

$$Z = \sum_j \exp\{-\beta E_j\} \tag{4}$$

$$= \sum_{n,m} \exp\{-\beta\mu h(n - m)\} \tag{5}$$

$$= \sum_n \frac{N!}{n!\,(N - n)!} \exp\{-\beta\mu hn\} \exp\{\beta\mu Hm\} \tag{6}$$

We start by the definition of the partition function, hence summing over all possible configuration j, each with its energy E_j, we then rewrite the summation in terms of the number of spins up n and down m in the system, and the energy E_j can be written as a function of n and m as well. The last step uses the fact that for a fixed number of lattice sites N we can write $m = N - n$ and then the binomial coefficient in front takes care of the number of configurations j that exist with a fixed number of up spins n.

Then we can use the binomial theorem to rewrite this as

$$Z = (\exp\{-\beta\mu h\} + \exp\{\beta\mu h\})^N \tag{7}$$

$$= 2\cosh^N(\beta\mu H) \tag{8}$$

So we have now found the partition function Z.

From the partition function we can get the energy of the system as

$$E = -\frac{\partial}{\partial\beta}\ln(Z) \tag{9}$$

$$= N\frac{\sinh\beta\mu H}{\cosh\beta\mu H}\mu H \tag{10}$$

One other property of the system that are we now able to describe is the Magnetization that is linked to the energy as follows

$$M = \frac{E}{N\mu H} \tag{11}$$

The magnetization represents the symmetry of the system, in the sense that it represents the relative difference of ups and downs $M =< \frac{n-m}{N} >$. From (11) we have that for the system we are describing the magnetization is

$$M = \frac{\sinh\beta\mu H}{\cosh\beta\mu H} = \tanh\beta\mu H = \tanh\frac{\mu H}{kT} \tag{12}$$

So the magnetization is a function of β and hence a function of temperature. For T big enough, hence β small enough, we have that the magnetization goes to 0 as we would expect. For high enough temperature indeed we would have that each point has equal probability to be in state up or down, hence on average the magnetization is 0.

We now consider the case where the interaction between near neighbours affects the energy of one point. We also do not look at the energy of a point itself but at the energy of a pair of points. For two points i and j

$$E_{ij} = -J\sigma_i\sigma_j \tag{13}$$

From this follows that the energy of the whole system can be rewritten as

$$E = -J\sum_n \sigma_n\sigma_{n+1} \tag{14}$$

The - sign in front of both E and E_{ij} is to favor the configuration where aligned spin have lower energy.

One other way to write this is instead of considering the values ±1 to be assigned to the points on the line we can assign a value ±1 to the piece in between them see figure 2. These pieces that connect the lattice sites are called edges.

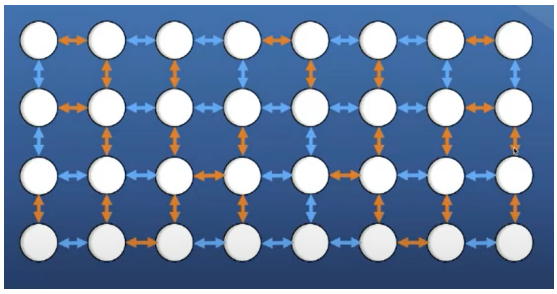

Figure 2: In this figure for a two-dimensional lattice, instead of assigning ± 1 to the lattice site themselves we are assigning values to the edges of the lattice, the edges indicate is to reach the lowest energy configuration the spins of the site should be aligned or not [mandi colloquia,]

In so doing we are simply considering the interaction between two neighbours. We would assign +1 in the case the spin of the two points are aligned and -1 in the opposite case. We can redefine the energy of a single edge or connection 'i' as

$$E_i = -J\tau_i \tag{15}$$

where τ_i is the value that describes the alignment or anti-alignment of the two spins, $\tau_i = \sigma_i \sigma_{i+1}$.

We can now calculate the partition function of this system where we have no external field but we only have that nearest neighbours interaction influence the total energy, keeping in mind that, in a one-dimensional lattice, for N points we have $N-1$ values of τ

$$Z = \sum_{i=1}^{2^{N-1}} \exp\{-\beta E_i\} \tag{16}$$

$$= \sum \exp\{-\beta J \sum_j \tau_j\} \tag{17}$$

$$= (2\cosh(\beta J))^{N-1} \tag{18}$$

Where the sum in equation (16) and the first sum in equation (17) is over all possible configuration i, the second sum in equation (17) is over all edges j of a given configuration i. The step in equation (18) follows from similar reasoning as the one in equation (5) and (6) by doing the summation over the 'up spin' 'n or parallel edges, and then as in equation (8) to find the final value of the partition function Z.

Now we can get the total energy of the system

$$E = -\frac{\partial}{\partial\beta}\ln(Z) \tag{19}$$

$$= -\frac{\partial}{\partial\beta}\left[(N-1)\ln(\cosh(\beta J))\right] \tag{20}$$

$$= -NJ\frac{\sinh\beta J}{\cosh\beta J} \tag{21}$$

$$= -NJ\tanh\beta J \tag{22}$$

From (14) we have that the average spin $<\tau>$ is

$$<\tau>=\frac{1}{N}\sum_i \tau_i = \tanh\beta J \tag{23}$$

3 Ising Model in Higher Dimensions

To be able to analyse more easily the higher dimensional case we can introduce the so-called Mean Field Approximation. The number of nearest neighbours in d dimensions for the square lattice of the cubic one and their respective in higher dimensions is 2d.

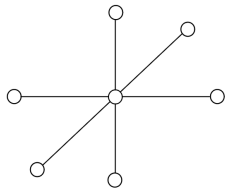

Figure 3: In the figure the case of the three-dimantional cubic lattice is taken .

If we then try to write the energy function for a single spin i as in figure 3 we have

$$E_i = -J\sigma_i\sum_k \sigma_k \tag{24}$$

here k are thee neighbours. The mean field approximation takes $\overline{\tau}$ to be the average τ. Where τ is again the representation of the interaction of two near spins $\tau_{ij} = \sigma_i\sigma_j$

$$<\sum_k \tau_k >= 2d\overline{\tau} \tag{25}$$

where $<\ >$ represents the mean of what is in the brackets. So we get for each spin i an energy

$$<E_i >= -J2d\sigma_i\overline{\tau} \tag{26}$$

If we calculate the partition function we get the similar value as we did in the one-dimensional case, and we get an average value for the spin we are considering, shown in figure 3.

$$< \sigma >= \tanh 2d\beta\tau \tag{27}$$

Since the model we are simplifying the system to has translation invariance we have that the average value of the spin we are considering, hence $< \sigma >$, has to be equal to the average of the other spins as well τ, this is the main implication of the Mean field approximation. This leads us to

$$< \sigma >= \tau = \tanh 2d\beta\tau \tag{28}$$

So we can now find a value for the average spin τ, we just have to solve the equation $\tau = \tanh 2d\beta\tau$.

We will solve this graphically but we first do a change of variables, using

$$Y = 2\beta dJ\tau \tag{29}$$

We now have to solve

$$\frac{Y}{2\beta dJ} = \tanh Y \tag{30}$$

Graphically we get figure 4

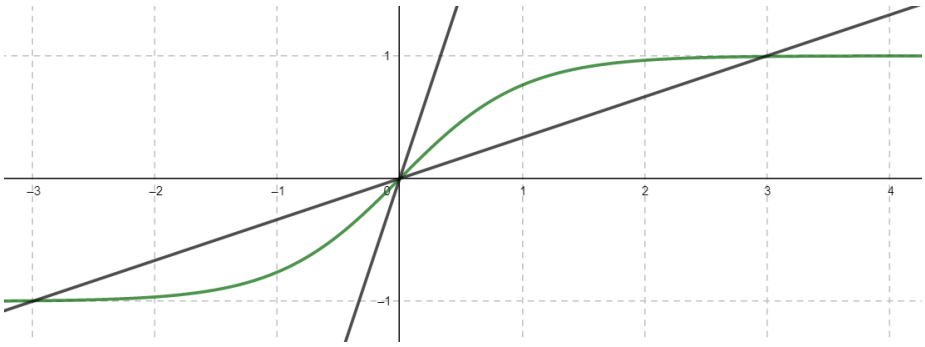

Figure 4: The picture shows units of Y on both axis. In green we have the hyperbolic tangent tanh, and in black we have two straight lines that go through the origin, the more vertical one indicates the high temperature scenario and the more horizontal one indicates the low temperature scenario with a non zero intersection point. [GmbH, 2002]

Where the coefficient of the line is $\frac{1}{2\beta dJ}$, but we know that β is just a function of the temperature $\beta = \frac{1}{Tk}$, so the coefficient of the line is actually a function of the temperature T. So we can rewrite the coefficient as $\frac{1k}{2dJ}T$.

For the temperature going to ∞ the coefficient hence goes to infinity so the lines tends to be vertical, for low temperature on the other hand the coefficient

goes to 0, hence the line is more horizontal. What we notice is that for certain values of this coefficient the line $y = \frac{kT}{2dJ}Y$ and the hyperbolic tangent cross once, high T value, or twice, low T value; we are only considering positive values of T, we are ignoring the solution we get at negative values of T.

To see for what value of T this behaviour is different we look at the slope of the hyperbolic tangent close to 0 and we have that this is 1 this means that for a coefficient $\frac{kT}{2dJ}$ with value 1 we have the change in behaviour; this happens exactly for $T = \frac{2dJ}{k}$ and this is the so called critical temperature T_c.

It is at this value of T that we have the phase transition; given that d, the dimensions of the model, higher or equal to 2.

What we have then is that for any value of T there is always a solution $Y = 0$. When $Y = 0$ we have that $Y = 2\beta dJ\tau = 0$, so we have that the average spin $\tau = 0$. As we have seen for values of Tk lower than $2dJ$ a second solution is possible, this solution gives us the average value of the spin of the system for a given temperature T

To see that this second one is actually the solution to our equation and not the Y =0 one, we start form the fact that at T = 0 we expect the system to be in the lowest energy state, such a state is one where all the spins are aligned and hence the average spin is $\tau = 1$.

To reconnect this to the graph we see that for $T \to 0$ we have two solutions for Y one which is 0 and the other one is non 0. From how we defined Y it follows that Y = 0 implies $\tau = 0$ so we can immediately conclude that for T going to 0 the Y = 0 can not be a physical solution, hence the other one must be. By continuity of the solution it follows that for all values of t less than T_c of the two possible solutions the non zero one is the physical one.

3.1 Small External Field

If we decide to introduce a small external magnetic field then we can redefine the energy as follow

$$E = -2dJ\tau\sigma - B\sigma \tag{31}$$

where the first part of the equation is the same as in the previous part and the second one is due to the interaction between the magnetic field B and the spin σ_i. In this case we can take $B << 1$ to be a perturbation. Following a similar derivation as the previous one for the partition function and then the average energy we can get tho the formula

$$\tau = \tanh\left(2dJ\beta\tau - \beta B\right) \tag{32}$$

If we solve graphically the equation we see that for a very small value of B we have a picture which is really close to figure 4, the main difference is that hyperbolic tangent is shifted to the left, it is no more centered at the origin.

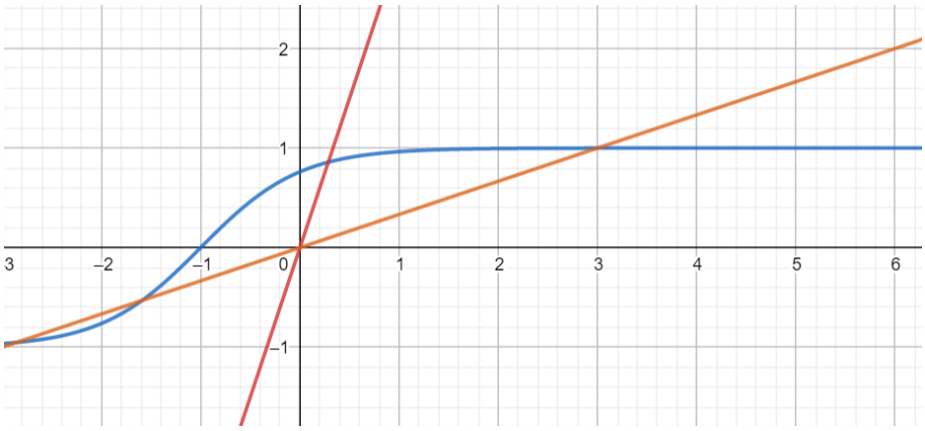

Figure 5: The picture shows units of Y on both axis. In blue we have the hyperbolic tangent moved to the left by a value taken to be constant. In red the high temperature line and in orange the low temperature one [GmbH, 2002].

Now in figure 5 there is only one solution, which is non zero, no matter the value of the temperature T. We can now check what happens for the two limit values of T going to infinity and to 0. In the first case as $T \to \infty$ we also have that $B\beta \to 0$, since $b\beta = \frac{B}{kT}$. So as the line $y = \frac{2dJ}{kT}$ tends to vertical the effects of the magnetic field tends to 0 and so the hyperbolic tangent goes back to being centered around the origin and we have the solution Y = 0 as in the previous case, which means that the average spin is 0. This again is what we expect for high temperature given that the probability of the system being in a particular state is equal at temperature going to infinity.

The case where $T \to 0$ we have the line $y = \frac{2dJ}{kT}$ going horizontal and at the same time $B\beta \to \infty$ so the hyperbolic tangent gets moved further to the left. The scenario where the temperature is low and we reach the lowest energy in the system, was a degenerate scenario in the absence of a magnetic field since the only thing that did matter was that the spins were aligned, either all up or all down. Now the presence of the smallest B field possible makes it such that one of these two configurations, that were at the same energy, is favored and the system is unstable. The tiniest perturbation in the magnetic field biases the system. This is what happens with a magnet, in absence of a magnetic field it has equal probability of being aligned in any direction, once it is brought to into the earth's magnetic field, for example, it aligns accordingly.

4 Mathematical Proof of the Phase Transition

Mathematical proof of phase transition in 2D or higher.

A more rigorous proof of the existence of the phase transition in two dimensions was made by Peierls [Peierls, 1936] in 1936. To show the existence of the phase transition what we have to do is show that for low temperature, hence

temperature T going to 0, we have a non-zero magnetization. An idea of how this can be done was given in a non mathematically rigorous way in the previous section.

So what we will show and what is proof of the phase transition is the existence of spontaneous magnetization, this is the tendency of the system to remain in the up state after a magnetic field has been applied and subsequently switched off.

The most common way, and the one used by Peierls as well, to make sure in a simulation that the external magnetic field is off is to impose the boundaries of the lattice we are considering to be all spin up, in this way we are assuming no external perturbation happening, since the boundaries are fixed and constant in time. We then let the boundary move to infinity.

We then look at the probability of the lattice pint at the origin to be in the 'down' state. If there were no external magnetic field the probability of this to be in either of the two possible states would be equal, hence our answer would be $\frac{1}{2}$.

What we will see for the low temperature limit, is that the effect of the past magnetic field, the imposed 'ups' at the boundary tend to influence the near sites and this effect propagates to all the lattice, the origin included. For high temperatures instead the effects quickly fade away and become negligible.

What we will show is that no matter the lattice size the probability of the origin to be in a 'down' state is lower than $\frac{1}{2}$, hence we see an effect of the 'past' magnetic field; this proofs that there is spontaneous magnetization, hence a phase transition for the 2D Ising model.

First of all recall that the probability of configuration ν is

$$P_\nu = \frac{\exp\{-\beta E_\nu\}}{Z} \tag{33}$$

where Z is the partition function and E_ν is the energy of the chosen configuration.

When we than have to calculate the probability for a specific lattice site σ_0 to be in the down state -1, we will sum over all configuration with such a site, hence

$$P_{\sigma_0=-1} = \frac{\sum_{\sigma\in\Omega_0}\exp\{-\beta E_\sigma\}}{Z} \tag{34}$$

where Ω_0 is the set of all configurations such that $\sigma_0 = -1$, as we want, and E_σ is the energy of each configuration.

For a given configuration the first step is to think of it as if we were to divide the 'ups' or '+' and 'downs' or '-' by a line as in figure 6

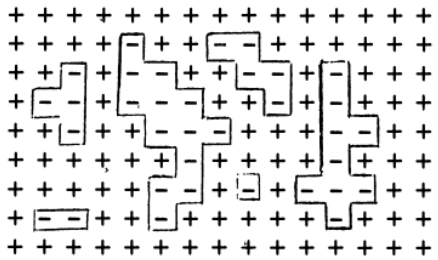

Figure 6: The figure shows Peierls representation and the way he separated the 'sea of pluses' and the 'islands of minuses' [Cipra, 1987]

Given that we impose the boundaries to be all ups, when Peierls describes this representation he talks about 'islands' of minus in a 'sea' of pluses 6. The lines separate the pluses and minuses, hence for each segment of such a line we have a bond $\sigma_i\sigma_j = -1$. These lines can be considered to be closed paths, hence start and end at the same point, but all other nodes are different from each other, so no nodes are touched twice.

Now if we pick all the configurations that have a minus sign at the origin, we have that there is such a line that encloses the origin, called this line S. We fix for now the length of this line and call it n.

The energy of each configuration can be written similarly as we have done in the previous sections as

$$E_\nu = -J \sum_{ij} \sigma_i \sigma_j \tag{35}$$

Again, this is the sum of all the interactions of nearest neighbours, where based on the '$-$' sign we have chosen the aligned configuration to be favoured. We can write the probability P_{Ω_S}, where Ω_s is the set of representations in Ω_0 where we have S as the line around the origin.

$$P_{\Omega_s} = \frac{1}{Z} \sum_{\sigma \in \Omega_S} \exp\{-\beta E_\sigma\} \tag{36}$$

$$= \frac{1}{Z} \sum_{\sigma \in \Omega_S} \exp\{-\beta \left[Jn(S) + \sum_{i,j \notin S} -J\sigma_i\sigma_j \right]\} \tag{37}$$

$$= \frac{1}{Z} \sum_{\sigma \in \Omega_S} \exp\{-\beta Jn(S)\} \quad \exp\{-\beta \sum_{i,j \notin S} -J\sigma_i\sigma_j\} \tag{38}$$

$$= \frac{1}{Z} \exp\{-\beta Jn(S)\} \sum_{\sigma \in \Omega_S} \exp\{-\beta \sum_{i,j \notin S} -J\sigma_i\sigma_j\} \tag{39}$$

Now for a fixed configuration in Ω_s we can create a new one by changing

all the spins inside the line S and keeping the rest fixed. We call this new configuration Ω'_S and the spins in this configuration σ'. We can see the following.

$$\sum_{i,j\in\Omega_s} \sigma'_i\sigma'_j = \sum_{i,j \text{ outside S}} \sigma'_i\sigma'_j + \sum_{i,j \text{ inside S}} \sigma'_i\sigma'_j + \sum_{i,j \text{ across S}} \sigma'_i\sigma'_j \tag{40}$$

Now we consider the three different summations we have with respect to the original representation in Ω_S. Given that in the mapping $\Omega_s \to \Omega'_s$ we only changed the spins inside the first summation gives us

$$\sum_{i,j \text{ outside S}} \sigma'_i\sigma'_j = \sum_{i,j \text{ outside S}} \sigma_i\sigma_j \tag{41}$$

If we now look at the spins inside S we have that each of them was flipped hence if two nearest neighbours were aligned they still are and if they were anti-aligned they still are so

$$\sum_{i,j \text{ inside S}} \sigma'_i\sigma'_j = \sum_{i,j \text{ inside S}} \sigma_i\sigma_j \tag{42}$$

For the spins across S we have that, by how S is defined each interaction was anti-aligned, but since we flipped only the spins on one side of S we now have that each spin pair is aligned so we have

$$\sum_{i,j \text{ across S}} \sigma'_i\sigma'_j = n(S) \tag{43}$$

The left-hand side of equation (38) and the right-hand side of equation (39) (40) (41) we get the equation

$$\sum_{i,j} \sigma'_i\sigma'_j = \sum_{i,j\notin S} \sigma_i\sigma_j + n(s) \tag{44}$$

$$\sum_{i,j\notin S} \sigma_i\sigma_j < \sum_{i,j} \sigma'_i\sigma'_j \tag{45}$$

Once we have this upper-bound for the interactions in the original representation we can go back to calculate the probability P_{Ω_s}

$$P_{\Omega_s} = \frac{1}{Z}\exp\{-\beta Jn(S)\} \sum_{\sigma\in\Omega_S} \exp\{-\beta \sum_{i,j\notin S} -J\sigma_i\sigma_j\} \tag{46}$$

$$< \frac{1}{Z}\exp\{-\beta Jn(S)\} \sum_{\sigma'\in\Omega'_S} \exp\{-\beta \sum_{i,j} -J\sigma'_i\sigma'_j\} \tag{47}$$

$$= \frac{1}{Z}\exp\{-\beta Jn(S)\} \sum_{\sigma'\in\Omega'_S} \exp\{-\beta E_{\sigma'}\} \tag{48}$$

In equation (45) we could replace the summation over $\sigma \in \Omega_S$ by $\sigma' \in \Omega'_S$ since se have that the mapping $\sigma \to \sigma'$ as we defined it before is one-to-one.

Now we use the inequality

$$\sum_{\sigma' \in \Omega'_S} \exp\{-\beta E_{\sigma'}\} = \sum_{\sigma'} \exp\{-\beta E_{\sigma'}\} \tag{49}$$

Since $\exp\{-\beta E_{\sigma'}\} > 0$ for any σ' we are simply adding positive terms to the summation.

We get

$$P_{\Omega_s} < \frac{1}{Z} \exp\{-\beta J n(S)\} \sum_{\sigma' \in \Omega'_S} \exp\{-\beta E_{\sigma'}\} \tag{50}$$

$$= \frac{1}{Z} \exp\{-\beta J n(S)\} \sum_{\sigma'} \exp\{-\beta E_{\sigma'}\} \tag{51}$$

$$= \frac{1}{Z} \exp\{-\beta J n(S)\} Z \tag{52}$$

$$= \exp\{-\beta J n(S)\} \tag{53}$$

Hence the final result

$$P_{\Omega_s} < \exp\{-\beta J n(S)\} \tag{54}$$

We now consider each possible line S to get

$$P_{\sigma_0 = -1} = \sum_S P_{\Omega_s} \tag{55}$$

$$< \sum_S \exp\{-\beta J n(S)\} \tag{56}$$

$$= \sum_{n=4} s(n) \exp\{-\beta J n\} \tag{57}$$

In equation (55) we are summing over the length of the line around the origin S. The value s(n) is the number of lines of length n.

The next step in the proof is to find an upper-bound for the number s(n). The way Peierls did it [Peierls, 1936] is not the best approximation possible but it works. Fist given n fixed we know that the line S cannot wander too far from the origin, and given that once it reaches a point it has to go back to where it started it needs to be inside a square with sides of length $n\sqrt{2}$, see figure 7.

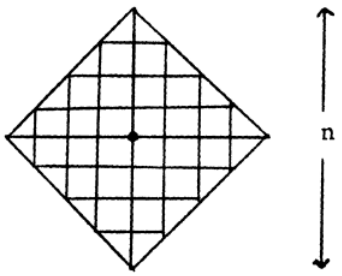

Figure 7: In the figure we have a square of side $\frac{\sqrt{2}n}{2}$ that includes all the sites that can be reached in $\frac{n}{2}$ and also includes all the closed paths of length n that start at the origin [Cipra, 1987]

We start by bounding the number of lines S of length n by using the number of closed random walks of length n $r(n)$.

$$s(n) < \frac{1}{n} r_{\text{closed}}(n) \tag{58}$$

The number of lines of length n is clearly less than the number of closed random walks with the same length, we can then add a factor of $\frac{1}{n}$ in front since each random walk is counted n times given that each of the point through which it goes can be considered to be the origin.

The next step is to consider all random walks not only the closed once and we have

$$s(n) < \frac{1}{n} r_{\text{closed}}(n) < \frac{1}{n} r(n) \tag{59}$$

The number of random walks of length n is $\frac{n^2}{2} 4^n$, so we have

$$s(n) < \frac{1}{n} \frac{n^2}{2} 4^n \tag{60}$$

$$= \frac{n}{2} 4^n \tag{61}$$

So back to the probability of having a minus sign at the origin.

$$P_{\sigma_0=-1} < \sum_{n=4} s(n) \exp\{-\beta J n\} \tag{62}$$

$$= \sum_{n=4} \frac{n}{2} 4^n \exp\{-\beta J n\} \tag{63}$$

$$= \frac{1}{2} \sum_{n=4} n(4 \exp\{-\beta J\})^n \tag{64}$$

$$< \frac{1}{2} \sum_{n=1} n(4 \exp\{-\beta J\})^n \tag{65}$$

Now to solve $\sum_{n=1} n(4 \exp\{-\beta J\})^n$ we use the fact that

$$\frac{1}{1-x} = \sum_{n=1} x^n \tag{66}$$

if we differentiate and multiply by x both sides we get t

$$\frac{x}{(1-x)^2} = \sum_{n=1} n x^n \tag{67}$$

We can use this to rewrite the sum we have

$$\sum_{n=1} n(4 \exp\{-\beta J\})^n = \frac{(4 \exp\{-\beta J\})}{[1 - (4 \exp\{-\beta J\})]^2} \tag{68}$$

So we conclude that

$$P_{\sigma_0=-1} < \frac{1}{2} \frac{(4 \exp\{-\beta J\})}{[1 - (4 \exp\{-\beta J\})]^2} \tag{69}$$

Since for β large enough, hence low enough temperatures we can make the right hand side arbitrarily small we conclude that for low temperatures

$$P_{\sigma_0=-1} < \frac{1}{2} \tag{70}$$

Hence we have spontaneous magnetization, hence a phase transition for the two dimensional Ising model on the square lattice no matter the size of the lattice.

This proof can be modified slightly and used not only for the two-dimensional square lattice but for other two-dimensional lattices as well.

Transfer matrix methods is the derivation Ising himself did and that shows no phase transition in the 1 dimensional case

5 Variation of the Ising Model

5.1 Bethe Lattice

So far we have only looked at the Ising model on the square lattice or higher dimensional versions of the square lattice. It is possible to study it with other lattice arrangements as well, one example is the so-called Bethe lattice. On this lattice configuration we are able to solve the Ising model, hence find a Critical temperature and the local magnetization under some assumptions.

To construct this lattice arrangement we start from a point we call 0, and we connect q lattice sites to the 0 point. These q points are the so-called 'first shell'. From each of these points we connect other $q-1$ points and so on. To each of the points in a shell s we connect $q-1$ points, only connected to them, and the set of all these points forms the $s+1$ shell.

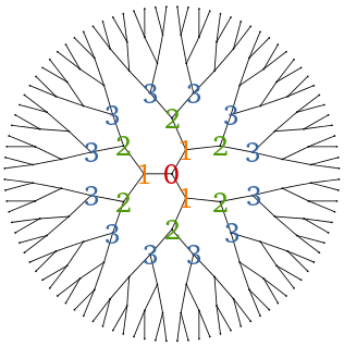

Figure 8: The picture represents a bethe lattice with coordination number $q = 3$, the site labelled with 0 would be what we called origin, and the sites belonging to the nth shell are labelled n [Fundation, 2015]

Given how we constructed our lattice all the points, besides the ones in the very last shell, are connected to q other sites.

In the s shell there are $(q-1)(\eta(s-1))$ points, where $\eta(s-1)$ is the number of lattices in the $s-1$ shell, hence we can rewrite this as $(q-1)(q-1)\eta(s-2)$, and so on remembering that in the first shell we have q lattice sites instead of $q-1$, hence in the s shell we have $q(q-1)^{s-1}$ sites in total.

In a lattice with q connections per site and a total of n shells we have a total

number of sites that is

$$\sum_{s=0}^{s=n} s(n) = 1 + \sum_{s=1}^{s=n} q(q-1)^{s-1} \tag{71}$$

$$= 1 + q \sum_{s=0}^{s=n-1} (q-1)^s \tag{72}$$

$$= 1 + q\frac{[1-(q-1)^n]}{1-(q-1)} \tag{73}$$

$$= 1 + \frac{q[(q-1)^n - 1]}{q-2} \tag{74}$$

This graph is called the Cayley tree. The outermost shell made of the sites that are only linked to one other point is the set of boundary points. In general when studying a system that has to resemble thermodynamic properties as we are doing, the ratio of number of boundary points over interior points should become small as the system becomes large enough, here $n \to \infty$. For the Cayley tree this is not the case, since both these numbers grow as q^n.

To avoid this problem we will consider only sites deep within the Cayley tree, hence far enough to the boundary points, these sites form the so-called Bethe lattice. If such an assumption is not made and one considers the partition function on the whole lattice, the contribution from the boundary points is non negligible and the system that is being studied is the Ising model in the Cayley tree, which is a system that has been solved as well.

The choice of the Bethe lattice is due to the fact that for a higher order approximation of the Critical temperature we need to know the number of triangles in the lattice, then the number of tetrahedrons, and so on with bigger polygons, this is easily avoided on the Bethe lattice since there are no closed paths.

Following the definition of dimensionality of a lattice [Baxter, 2016b] we have that from

$$c_m = 1 + m_1 + m_2 + m_3 + ... \tag{75}$$

where m_i is the number of lattice site distant i from a starting point, then the dimensionality of a lattice is d where

$$d = \lim_{n\to\infty} \frac{\ln(c_n)}{\ln(n)} \tag{76}$$

In the appendix we can see how this works for some cases of regular 2 and 3 dimensional lattices.

For the bethe lattice the number c_n is given in equation (73) and we have $c_m \sim \frac{q[(q-1)^n-1]}{q-2}$.

$$\lim_{n\to\infty}\frac{\ln(c_m)}{\ln(n)} \sim \lim_{n\to\infty}\frac{\ln\left(\frac{q[(q-1)^n-1]}{q-2}\right)}{\ln(n)} \tag{77}$$

$$\sim \lim_{n\to\infty}\frac{\ln\left(\frac{q(q-1)^n}{q-2}\right)}{\ln(n)} \tag{78}$$

$$\sim \lim_{n\to\infty}\frac{\ln\left((q-1)^n\right)}{\ln(n)} \tag{79}$$

$$\sim \lim_{n\to\infty}\frac{n\ln(q-1)}{\ln(n)} \tag{80}$$

$$\to \infty \tag{81}$$

hence following this definition the Bethe lattice is an infinite dimensional lattice.

5.2 Finding the Magnetization

We start by writing the complete partition function for a system on the Cayley lattice, we will later ignore the boundary terms, with only nearest neighbours interactions and an external magnetic field non zero. For the above assumptions we have that the partition function is given by

$$Z = \sum_{\sigma}\exp\left\{K\sum_{i,j}\sigma_i\sigma_j + h\sum_i\sigma_i\right\} \tag{82}$$

The first summation is over each pair of directly connected sites i, j, or each edge of the graph, the second summation is over each site i, h represents the external magnetic field and K represents the interaction between the pairs and we have $K = J\beta = \frac{J}{kT}$.

We want to find the average magnetization of a site in the bethe lattice, so far enough from the boundary points. The Magnetization is by definition the average of the spin so

$$M =< \sigma_0 >= \frac{1}{Z}\sum_{\sigma}\sigma_0 P(\sigma) \tag{83}$$

So we are summing over all possible configurations σ.

$$P(\sigma) = \exp\left\{K\sum_{i,j}\sigma_i\sigma_j h\sum_i\sigma_i\right\} \tag{84}$$

the probability of each configuration σ.

If we think at the graph as disconnected at the origin' 0 it looks like it is made of q identical disconnected graphs. We can then rewrite the probability

$P(\sigma)$ by thinking of it as the product of the probabilities of each of these graphs hence

$$P(\sigma) = \exp\{h\sigma_0\} \prod_{j=1}^{q} Q_n(\sigma_0|s_j) \tag{85}$$

Where s_j denotes the configuration of spins of one disconnected graph, same way σ did for the whole graph, σ_0 is the spin of the site 0, and $Q_n(\sigma_0|s_j)$ is the function

$$Q_n(\sigma_0|s_j) = \exp\{K \sum_{i,j} s_i s_j + K s_1 \sigma_0 + h \sum_i s_i\} \tag{86}$$

The first summation is over all the edges in the sub-graph, beside the one with the 0 site (0,1), already taken into account by the term $Ks_1\sigma_0$, is the interaction with the magnetic field. The subscript n indicates the number of shells the sub-graph has.

If we then look at one of the q subtrees we have created we can do the same thing we have done with the original tree and 'disconnect' the point 1 so that we have other q identical sub-sub-trees. This then allows us to redefine $Q_n(\sigma_0|s_j)$.

$$Q_n(\sigma_0|s_j) = \exp\{K\sigma_0 s_1 + h s_1\} \prod_{j=1}^{q-1} Q_{n-1}(s_1|t_j) \tag{87}$$

where t_j has the same role s_j had in $Q_n(\sigma_0|s_j)$, hence it denotes the configuration of all the spins in the j^{th} sub-sub-tree.

These definitions of $P(\sigma)$ and $Q_n(\sigma_0|s_j)$, make it easier to calculate M.

Now we define

$$g_n(\sigma_0) = \sum_s Q_n(\sigma_0|s) \tag{88}$$

So now we can rewrite the partition function as

$$Z = \sum_\sigma P(\sigma) \tag{89}$$

$$= \sum_\sigma \exp\{h\sigma_0\} \prod_{j=1}^{q} Q_n(\sigma_0|s_j) \tag{90}$$

$$= \sum_{\sigma_0} \exp\{h\sigma_0\} \prod_{j=1}^{q} \sum_{\sigma \setminus \sigma_0} Q_n(\sigma_0|s_j) \tag{91}$$

$$= \sum_{\sigma_0} \exp\{h\sigma_0\} \prod_{j=1}^{q} g_n(\sigma_0) \tag{92}$$

$$= \sum_{\sigma_0} exp\{h\sigma_0\} \, [g_n(\sigma_0)]^q \tag{93}$$

So from equation (82) it follows that

$$M = \frac{1}{Z}\sum_{\sigma_0} \sigma_0 \exp\{h\sigma_0\}[g_n(\sigma_0)]^q \tag{94}$$

Now we define

$$x_n = \frac{g_n(\sigma_0 = -1)}{g_n(\sigma_0 = +1} \tag{95}$$

or in short $x_n = \frac{g_n(-)}{g_n(+)}$. We then have anew formula for M

$$M = \frac{1}{Z}\sum_{\sigma_0} \sigma_0 \exp\{h\sigma_0\}[g_n(\sigma_0)]^q \tag{96}$$

$$= \frac{(+1)\exp\{h(+1)\}[g_n(+)]^q + (-1)\exp\{h(-1)\}[g_n(-)]^q}{exp\{h(+1)\}[g_n(+)]^q + \exp\{h(-1)\}[g_n(-)]^q} \tag{97}$$

$$= \frac{exp\{h\} - \exp\{-h\}x_n^q}{exp\{h\} + \exp\{-h\}x_n^q} \tag{98}$$

To get M we only need to know x_n. Take equation (85) and sum over s_j on both sides, then

$$\sum_{s_j} Q_n(\sigma_0|s_j) = \sum_{s_j} exp\{K\sigma_0 s_1 + hs_1\}\prod_{j=1}^{q-1} Q_{n-1}(s_1|t_j) \tag{99}$$

$$g_n(\sigma_0) = \sum_{s_1} \exp\{K\sigma_0 s_1 + hs_1\}\prod_{j=1}^{q-1}\sum_{s\setminus s_1} Q_{n-1}(s_1|t_j) \tag{100}$$

$$g_n(\sigma_0) = \sum_{s_1} \exp\{K\sigma_0 s_1 + hs_1\}[g_{n-1}(s_1)]^{q-1} \tag{101}$$

We expand the summation over s_1 and we get

$$\frac{g_n(-)}{g_n(+)} = \frac{exp\{K(-1)(+1) + h(+1)\}[g_{n-1}(+)]^{q-1} + \exp\{K(-1)(-1) + h(-1)\}[g_{n-1}(-)]^{q-1}}{exp\{K(+1)(+1) + h(+1)\}[g_{n-1}]^{q-1} + \exp\{K(+1)(-1) + h(-1)\}[g_{n-1}(-)]^{q-1}]} \tag{102}$$

$$x_n = \frac{exp\{-K + h\}[g_{n-1}(+)]^{q-1} + \exp\{K - h\}[g_{n-1}(-)]^{q-1}}{exp\{K + h\}[g_{n-1}(+)]^{q-1} + \exp\{-K - h\}[g_{n-1}(-)]^{q-1}} \tag{103}$$

$$x_n = \frac{exp\{2h\} + \exp\{2K\}[x_{n-1}]^{q-1}}{exp\{2K + 2h\} + [x_{n-1}]^{q-1}} \tag{104}$$

So now we have this recursive formula for x_n, and if we have x_0, we can then get the values for all the x_n, and then the value of M. We can see that $x_0 = g_0(\sigma_0) = 1$.

5.3 Limit as $n \to \infty$

Now that we have this expression for the magnetization we can let $n \to \infty$ and in so doing we are moving the boundary terms, the n^{th} shell infinitely far away from the points we are considering that we get the Magnetization per site of points deep enough in the Cayley three, hence points of the Bethe lattice.

The behaviour of x_n as $n \to \infty$ is determined by how many times the lines $y = x$ and $y = \frac{exp\{2h\}+\exp\{2K\}x^{q-1}}{exp\{2K+2h\}+x^{q-1}}$ from equation (104).

Using Descartes rule for positive real solution of a polynomial, see the link in the references [Wikipedia contributors, 2022], we can see that the equation (104) for x the limit of x_n when n goes to ∞ can have either 1 or three solution

In the case there is only one solution, for $n \to \infty$ we have that the limit will always monotonically approach the intersection point.

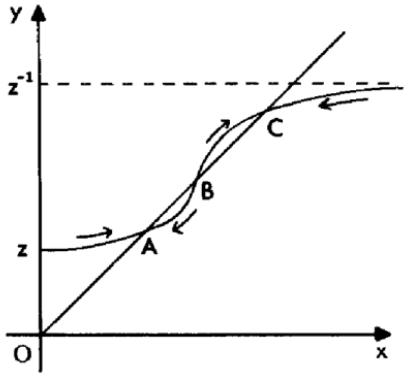

Figure 9: This figure shows that in the case we have three possible intersections the solutions converge to: A in the case of the starting value of x being smaller than B and to C if it is bigger than B[Baxter, 2016a].

The other possibility is that the lines intersect three times, as in figure 9, in this case we have two stable limit point and an unstable one. The unstable one is the one in the middle; solution below the first point, A in the figure, or above the last one, C in the figure, converge monotonically to it. Solution in between these two points, approach the closest one in the limit. The only time we have the point B as a limit in when we have $x = y(x) = 1$ and this happens only for $h = 0$, for $h < 0$ we have that the solution converges to A and for $h > 0$ it converges to C.

5.4 Critical Temperature

We now try to find the maximum value our x can take and this is the value of the critical temperature of this system.

From equation (104) we can obtain the following

$$\exp\{2h\} = x^{q-1}(\exp\{2K\} - x)(\exp\{2K\}x - 1) \tag{105}$$

given that all the x_n are positive, so their limit x will be as well and that the left-hand-side of the equation is positive and the right one must be as well, we have that the value of x lies in the interval $\exp\{-2K\} < x < \exp\{2K\}$.

Now we differentiate equation (105), instead of directly differentiating both sides of the equation we first take the logarithm of both sides and then differentiate, since the logarithm is a monotonically increasing function the two maximum will be the same.

$$\frac{d}{dx}\ln\left(\exp\{2h\}\right) = \frac{d}{dx}\ln\left(x^{q-1}(\exp\{2K\} - x)(\exp\{2K\}x - 1)\right) \tag{106}$$

$$2\frac{d}{dx}h = \frac{q-1}{x} - \frac{1}{\exp\{2K\} - x} - \frac{\exp\{2K\}}{\exp\{2K\}x - 1} \tag{107}$$

$$2\frac{d}{dx}h \cdot x = (q-1) - x\left[\frac{\exp\{2K\}x - 1 + \exp\{4K\} - \exp\{2K\}x}{\exp\{2K\}\exp\{2K\}x - \exp\{2K\} - \exp\{2K\}x^2 + x}\right] \tag{108}$$

$$2\frac{d}{dx}h = (q-1) - x\left[\frac{\exp\{4K\} - 1}{\exp\{4K\}x - \exp\{2K\} - \exp\{2K\}x^2 + x}\right] \tag{109}$$

$$2\frac{d}{dx}h = (q-1) - \frac{2\sinh(2K)}{\exp\{2K\} - \frac{1}{x} - x + \exp\{-2K\}} \tag{110}$$

$$2\frac{d}{dx}h = (q-1) - \frac{2\sinh(2K)}{2\cosh(2K) - x - \frac{1}{x}} \tag{111}$$

where in equation (110) we used the identity

$$2\sinh(2K)\frac{1}{\exp\{2K\}} = \exp\{4K\} - 1 \tag{112}$$

and in equation (111) we used

$$2\cosh 2K = \exp\{2K\} + \exp\{-2K\} \tag{113}$$

For x in the given interval we see that the function attains its maximum for $x = 1$ And based on the value of K, we have a different behaviour of the equation (104). The critical value is $K = \frac{1}{2}\ln(\frac{q}{q-2})$ which gives us a value of the critical temperature as

$$K_C = \frac{J}{kT_C} = \frac{1}{2}\ln(\frac{q}{q-2}) \tag{114}$$

for a deeper discussion of the behaviour of the equation as we vary h see Exactly Solved Models in Statistical Mechanics by R.J. Baxter [Baxter, 2016a].

5.5 Random Cluster Model

The Ising model has been studied for many decades now and from it several other models were derived. One in particular links percolation theory, the Ising model and the Potts model, and has borne much progress to this field of statistical physics, the Random Cluster model.

When describing this model we start by considering a graph $G = (V, E)$, where V is the set of vertices on the graph and E is the set of edges. An edge $e \in E$ is indicated by $e =< x, y >$ and is a link between vertices $x, y \in V$.

In the random Cluster Model we speak of edges being either open or close. Given a probability measure in the set of subsets of E, and the states space $\Omega = \{0, 1\}^E$, we call and edge open if $\omega(e) = 1$, and closed if $\omega(e) = 0$; where $\omega \in \Omega$ are vectors of zeros and ones.

Taking $\nu(\omega) = \{e \in E : \omega(e) = 1\}$, we see that there is a one to one correspondence between the vectors ω and the subsets F of E, $F \subset E$; given by $F = \nu(\omega)$

We also define $k(\omega)$ to be the number of connected components for a given configuration ω. These connected component or open clusters include isolated vertices.

Once we have defined all of the above we have that a random cluster measure has two parameters $q \in (0, infty)$ and $p \in [0, 1]$, and is defined as

$$\phi_{p,q}(\omega) = \frac{1}{Z}\left[\prod_{e\in E} p^{\omega(e)}(1-p)^{1-\omega(e)}]\right] q^{k(\omega)} \tag{115}$$

where again Z is the partition function

$$Z = \sum_{\omega\in\Omega}\left[\prod_{e\in E} p^{\omega(e)}(1-p)^{1-\omega(e)}]\right] q^{k(\omega)} \tag{116}$$

For specific values of q we can reconnect the Random Cluster Model to models that are already known and studied. Given the measure in equation (115) the first thing we see is that based no the value of q the measure favours or disfavours the formation of separated components, for $q < 1$, disconnected components are disfavoured, while for $q > 1$ they are favoured. For $q = 1$, the edges are open or closed independently of the other and this special case is the case of Percolation and random graph.

In the study of this model we focus on integer values of q only, given that the non-integer case is harder to study and seems to be less 'important' in its application.

For $q = \{2, 3, 4, 5...\}$ we have that the random cluster model is the same as the Potts model. This model is a generalization of the Ising model in the sense that while the Ising model allows only two spin orientation, 'up' or 'down' or ± 1, the Potts model allows for $q = \{2, 3, 4, 5, ...\}$ possible spin values, so the Potts model with $q = 2$, simply is the Ising model.

What we will do to show the connection between the Random cluster model and the Potts model, hence the Ising model as well, is to define the probability

measure of both these models, and show that there is a coupling between the two; then we will also show how to get from the representation of one model to the other.

The probability measure of the random cluster model is already given in equation (115), the one of the pots model is given as follow

$$\pi_{\beta,q}(\sigma) = \frac{1}{Z}\exp\{\beta H(\sigma)\} \tag{117}$$

with $Z = \sum_{\sigma}\exp\{\beta H(\sigma)\}$ and where H is the Hamiltonian and it is similar to that of the Ising model

$$H = \sum_{e=<x,y>\in E} -\delta_e(\sigma) \tag{118}$$

Where $\delta_e(\sigma) = \delta_{\sigma_x,\sigma_y}(\sigma)$ is the kronecker delta and is equal to 1 if the vertices of the edge e have the same spin orientation.

To switch from the potts measure to the Ising one we can use the identity

$$\delta_{\sigma_x,\sigma_y}(\sigma) = \frac{1}{2}(1+\sigma_x\sigma_y) \tag{119}$$

If σ_i can only take two values, hence $q = 2$ as in the case of the Ising model.

we can see this by doing the substitution and knowing from the previous paragraphs that we should get a probability measure for the Ising model of the form

$$\pi_{\beta}(\sigma)\frac{1}{Z}\exp\left\{\beta\sum_{e=<x,y>\in E}\sigma_x\sigma_y\right\} \tag{120}$$

If we substitute we get that the Hamiltonian of the Ising model H_I is

$$H_I = \sum_{e=<x,y>\in E} -\delta_e(\sigma) \tag{121}$$

$$= \sum_{e=<x,y>\in E} -\frac{1}{2}(1+\sigma_x\sigma_y \tag{122}$$

$$= -\frac{1}{2}|E| - \frac{1}{2}\sum_{e=<x,y>\in E}\sigma_x\sigma_y \tag{123}$$

Given this expression for H_I we have that the probability measure is from (117)

$$\pi_{\beta,q=2}(\sigma) = \frac{1}{Z}\exp\{\beta H_I(\sigma)\} \tag{124}$$

$$= \frac{1}{\sum_\sigma \exp\Big\{\beta[\frac{1}{2}|E|+\frac{1}{2}\sum_{e=<x,y>\in E}\sigma_x\sigma_y]\Big\}}\exp\left\{\beta[\frac{1}{2}|E|+\frac{1}{2}\sum_{e=<x,y>\in E}\sigma_x\sigma_y]\right\} \tag{125}$$

$$= \frac{\exp\{-\frac{1}{2}\beta|E|\}}{\sum_\sigma \exp\Big\{\frac{1}{2}\beta\sum_{e=<x,y>\in E}\sigma_x\sigma_y\Big\}}\exp\left\{\frac{1}{2}\beta|E|\right\}\exp\left\{\frac{1}{2}\beta\sum_{e=<x,y>\in E}\sigma_x\sigma_y\right\} \tag{126}$$

$$= \frac{1}{\sum_\sigma \exp\Big\{\frac{1}{2}\beta\sum_{e=<x,y>\in E}\sigma_x\sigma_y\Big\}}\exp\left\{\frac{1}{2}\beta\sum_{e=<x,y>\in E}\sigma_x\sigma_y\right\} \tag{127}$$

$$= \frac{1}{Z_I}\exp\left\{\frac{1}{2}\beta\sum_{e=<x,y>\in E}\sigma_x\sigma_y\right\} \tag{128}$$

Now what we have shown that the Pots model for $q = 2$ has the same probability measure we con show the link between the Potts model and the Random cluster model and from this it will follow the one with the Ising model as well.

We define a probability measure common to both model on the product space $\Sigma \times \Omega$, where $\Sigma = \{2, 3, 4, ..., q\}^V$, $\Omega = \{0, 1\}^E$.

We then define the probability mass function μ on this space as

$$\mu(\sigma, \omega) \propto \prod_{e\in E}\left\{(1-p)\delta_{\omega(e),0} + p\delta_{\omega(e),1}\delta_e(\sigma)\right\} \tag{129}$$

where again $\delta_e(\sigma) = \delta_{\sigma_x\sigma_y}$. Then we have to have a constant in front of the probability mass function in order for it to be normalized.

About this probability measure we can say the following facts: Assume $p = 1 - \exp\{-\beta\}$

- The marginal measure that we call $\mu_1(\sigma) = \sum_{\omega\in\Omega}\mu(\sigma, \omega)$ on Σ is the Potts measure
$$\mu_1(\sigma) = \frac{1}{Z_P}\exp\left\{\beta\sum_{e\in E}\delta_e(\sigma\right\} \tag{130}$$
where $\frac{1}{Z_P}$ is the normalization constant
- The marginal measure $\mu_2(\omega) = \sum_{\sigma\in\Sigma}\mu(\sigma, \omega)$ on Ω is the random-cluster measure
$$\mu_2(\omega) = \frac{1}{R_{RC}}\left\{\prod_{e\in E} p^{\omega(e)}(1-p)^{1-\omega(e)}\right\}q^{k(\omega)} \tag{131}$$

where again $\frac{1}{Z_{RC}}$ is the normalization constant

- We have

$$\sum_{\omega\in\Omega}\mu_1 = \sum_{\sigma\in\Sigma}\mu_2 \tag{132}$$

$$\sum_{\omega\in\Omega}\left\{\prod_{e\in E} p^{\omega(e)}(1-p)^{1-\omega(e)}\right\} q^{k(\omega)} = \sum_{\sigma\in\Sigma}\prod_{e\in E}\exp\{[\beta(\delta_e(\sigma)-1)]\} \tag{133}$$

$$\tag{134}$$

which gives us

$$Z_{RC}(p,q) = \exp\{-\beta|E|\}Z_P(\beta,q) \tag{135}$$

This were the marginal measures of μ and we have that the conditional measures assuming $p = 1-\exp\{-\beta\}$ are

- $\mu(.|\omega)$ is obtained starting from a given cluster configuration and by randomly putting spin on the given lattice. These spins are taken to be constant over the same cluster and the distribution that these random spin follow is uniform on $\{1,2,...,q\}$
- the measure $\mu(.|\sigma)$ is obtained by starting from a lattice with given spin per lattice site and for the edge $e =< x,y >$ we have that $\sigma_x \neq \sigma_y$ then we set $\omega(e) = 0$ so the edge is not part of the model, otherwise if $\sigma_x = \sigma_y$ we have that $\omega(e) = 1$ with probability p and 0 otherwise

We have then defined a measure μ that is a coupling of the Potts measure $\pi_{\beta,q}$ and the random cluster measure $\phi_{p,q}$ and the parameters of these relation are linked by the identity $p = 1-\exp\{-\beta\}$.

6 Spin Glasses

In this second part of this thesis we will see the application of the Ising model to the spin glasses, and then further application of these physical systems.

The name spin glasses come from the disordered nature of these systems. In the sense that spin glasses are the magnetic equivalent of structural glasses. What in physics is generally meant by structural glass is a solid in which the position of the molecules, or atoms, in space is not ordered but random. With spin glasses it is not the position of the atoms that is disordered but rather the direction of the spin of atoms arranged in a regular lattice, see figure 10.

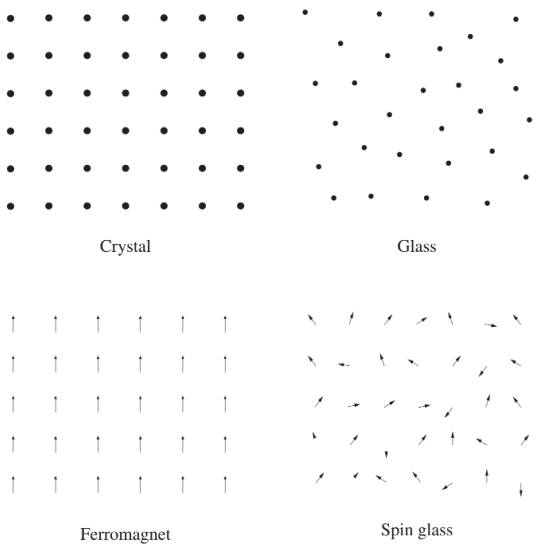

Figure 10: In this figure we show the difference between a crystal where the atoms are arranged in a regular lattice and a structural glass where the atoms present disorder in their spatial arrangement. This is compared with the case of a ferromagnet, where the spins are aligned, and a spin glass that presents disorder in the spin orientation, [Stein and Newman, 2013].

This disorder is not thermal disorder, given by a high energy in the system, but it is quenched disorder that persists at low temperature.

The term quench derives from metallurgy where it refers to plunging a metallic material like steel at a very high temperature into a cold liquid, like water. The reason behind this is to preserve some properties that might be lost under a slower cooling. Fast cooling, the same way it happens with structural glasses, like the one in our windows, does not allow the atoms or molecules in the material to arrange in an ordered manner; this slower way of cooling is referred to as annealing, which will be explained more later.

The first systems studied under the name of spin glasses were an alloy usually

of gold and traced of iron, that with a specific composition of the two metals presented a sharp peak in the magnetic susceptibility of the alloy indicating a phase transition. When this phase transition was then compared to the 'normal' state phase transition the heat peak is not found where the one of the magnetic susceptibility would suggest. The magnetic properties of these first materials were studied and from this the field of spin glasses has developed and broadened to many different fields of physics.

Spin glass systems present peculiar properties that are worth investigating. One of these is the fact that once they are in the so-called spin glass phase they are 'stuck' in a certain spin configuration, this is sort of what happens with structural glasses where the position of the molecules is somewhat constant on human time scales. This spin glass phase can be studied and it is detected by looking at the long time average spin of each lattice site. In this thesis we will not focus on this aspect of these systems but mainly on their link to combinatorial optimization problems, but a slightly more advanced explanation of this phase can be found in this video [mandi colloquia, 2021].

6.1 EA Model and Frustration

To describe the behaviour of spin glasses Edwards and Anderson redefined the Ising model by having the variable J to be a random variable and not a constant, so the model of the spin glasses can be described by an Hamiltonian of the form

$$H = -\sum_{<i,j>} J_{i,j}\sigma_i\sigma_j \tag{136}$$

and $J_{i,j}$ follows the probability distribution that usually has a mean that goes to 0 ans the number of lattice sites N goes to infinity, and variance that scales as $\frac{1}{\sqrt{N}}$, so a gaussian distribution can do the trick

$$P(J_{i,j}) = \left(\frac{N}{2\pi J^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{N}{2J^2}\left(J_{ij} - \frac{j_0}{N}\right)^2\right\} \tag{137}$$

for constants J and J_0 [Edwards and Anderson, 1975].

The main difference with the Ising model is that instead of having a fixed value for J in the Hamiltonian we now have a random variable J_{ij}, moreover we have that this variable can be both positive and negative.

What this implies is that for given configurations we have that there are bonds that cannot be satisfied no matter the spin arrangements, this phenomenon is called frustration, see figure 11.

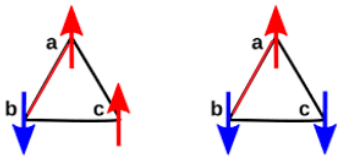

Figure 11: This figure is a representation of the phenomenon of frustration. the red edges in the figure represent a negative value of the interaction strength $J_{i,j}$ and the black one a positive value. If we look at the left figure we see that the bottom edge is not satisfied while the other two are, now no matter which spin we flip we will never be able to satisfy all the bonds [Gualtieri,].

One simple example of frustration is shown in figure 11 where no matter how we flip the spin at the bottom right corner of the square there is always a bond that is not satisfied.

Apart from frustration one other difference that we have in the Edwards-Anderson model is that instead of having a doubly degenerate ground state (all spins up or all spins down in the ferromagnetic Ising model) the degeneracy is higher and there are many states with an energy level very close to that of the ground state.

The high number of states close in energy and the fact that these states may differ a lot one another is the cause of the characteristic energy landscape of the spin glass systems. Flipping a spin may cause a high increase in the energy of the system, but after many flips one can reach a state with an energy close or equal to the original one; the landscape of these systems is characterized by a high number of mountains and valleys, that will be an essential characteristic for one of the application of spin glasses .

Frustration is then the cause of the characterised energy landscape that presents many local minima close in energy, as we can see in figure 12, and the global minimum does not generally present a big energy difference with the first excited states, and this makes it hard to identify it.

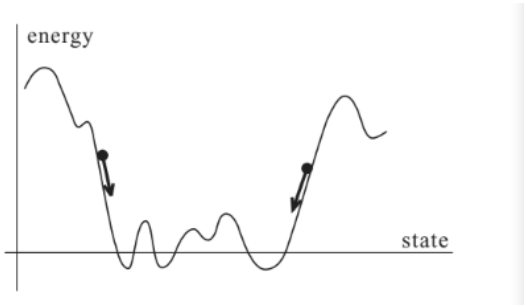

Figure 12: This figure shows the characteristic landscape of spin glass systems, where we have many local minima close in energy to the global one, that represents the ground state of the system [Nishimori, 2001].

6.2 Free energy

One characteristic of the system that we have not considered too much throughout the thesis is the Free energy of the system. Finding the free energy of a system is one of the goals when studying it since from it we can then derive the partition function and the magnetization, hence the critical temperature.

The free energy f is expressed as

$$f = -kT \ln (Z) \tag{138}$$

where k is the Boltzmann constant, t is the temperature and Z is the partition function.

One more thing about the free energy is that in the thermodynamic limit, hence for the number of lattice site N going to infinity, we expect that the free energy is proportional to the size of the system, hence the expression

$$f = -kT \lim_{n \to \infty} \frac{1}{N} \ln Z \tag{139}$$

has a finite value, so the limit exists and is non zero. Such quantities are so called self averaging. For low range interaction this characteristic of the free energy can be proved see [Castellani and Cavagna, 2005], but for infinite range interaction system this has not been proven yet, but numerical computation of the free energy clearly show that this property holds in these cases as well.

In models where the interaction strength J is a random variable instead of looking at the value of f we look at its average value. We do so since the non-averaged value of f would be dependent on the random realization of the values J_{ij} which would imply that for the same model we get different result based on these realizations; the average on the other hand will be the same no matter the realization of the interaction strength.

6.3 Replica Trick

Now as we said to understand the thermodynamic properties of the system we want to find the average of the free energy, but this would require us to calculate the average of equation (138) The problem here is that calculating the average of the logarithm of the partition function $\overline{\ln Z}$ is quite difficult.

To try to solve this problem, a known equality was used already in the first paper by Edwards and Anderson [Edwards and Anderson, 1975].

$$\overline{\ln(Z)} = \lim_{n\to\infty} \frac{\overline{Z^n} - 1}{n} \tag{140}$$

What Parisi did ad that gave us the possibility to be able to better understand these kind of problem and it is worthy of half of the Nobel prize of 2021, is to interpreter $\overline{Z^n}$ as the average of the partition function of n different system each with its different configuration of the random variable $J_{i,j}$. Now once we have this idea in mind computing the average of the products of these partition function is much simpler and a further calculation and explanation is also given in the paper [Guerra, 2004], or Parisi and others book [Mézard et al., 1987].

7 Combinatorial Optimization Problem

Combinatorial optimization problems are generally speaking problems where the goal is to find the best solution among many possible ones, in other words to find the minimum in the cost function of the system.

One example of such problems is the so-called Travelling Salesman Problem (TSP) where the goal is to find the shortest closed path that connects a given set of points.

7.1 P and NP Problems

These kinds of problems are still one of the biggest challenges in computational sciences, since they belong to the so-called NP problem.

In short the main distinction in computational complexity theory is between P and NP problems. The former are all those problems that using the best known algorithm can be solved in polynomial time, the latter are those that cannot be solved in polynomial time. This is an oversimplification of a more complex problem, but it is enough for what we need to treat.

What this means is that for NP problems the issue arises when we deal with large N, where N is for example the number of points we have to go through in the Travelling Salesman Problem; note that $N = 50$ is already big enough. The problem is that the time modern, classical, computers take to solve these kinds of problems increases non-polinomially with this number N, which makes it impossible for us, at this moment, to solve such problems in a reasonable amount of time.

There are two more sub-classes of NP problems which are worth discussing and these are the NP-complete and NP-hard problems, and most combinatorial

optimization problems belong to these classes. The former is a subset of the NP problem and each of the problem belonging to this class can be translated into any other NP problem, the NP-hard problem on the other hand are at least as hard to solve as any other NP problem, they are not a subset of the NP problem but they can as well be translated into any other NP problem, a visualization of these classes is shown in figure 13.

What is important to remember about these two classes is that being able to solve one of these problems would mean being able to solve all of the NP-problems.

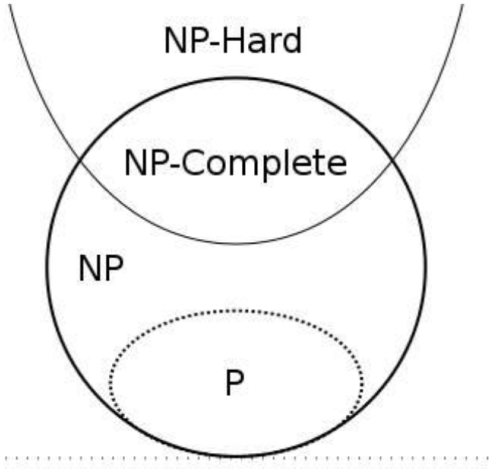

Figure 13: This figure gives a representation of how P, NP and other types of Np problem are related to one another. The displayed representation is thought to have increasing complexity the higher we are, so NP-hard are the hardest problem to solve and P problems are the easiest one [Sürmeli,].

It is known that the cost function of combinatorial optimization problem can be expressed as the Hamiltonian of the corresponding random Ising model; so it is possible to map one problem to a random Ising model. Once this is done, if we find the ground state of the given Hamiltonian we have found the best solution to our optimization problem and there are algorithms that can be used to solve these systems.

7.2 Travelling Salesman Problem

We will now work out an example on how we can map one combinatorial optimization problem as the travelling salesman problem to the Ising model.

The goal of this problem is to find the minimum path under the following conditions

- there are N cities

- we call the distance between two cities i and j, l_{ij}
- the salesman can go through a city only once
- the salesman goes back to the initial city after it has gone through all the others.

We can then write the cost function of this problem, hence what we try to minimize as

$$H = \sum_{\alpha=1}^{N} l_{c_\alpha, c_{\alpha+1}} \tag{141}$$

where c_α is the city at the αth step; and the last condition can be written as $c_{N+1} = c_1$. Again our goal is to find the set $\{c_1, c_2, ..., c_N\}$. When the number N is small enough this can be done by brute force, hence comparing all possible solutions, but fr N large this is impossible with classic computers since the number of solution scales as

$$\frac{(N-1)!}{2} \tag{142}$$

which is non polynomial, hence this is an NP problem.

In this as another problem we have to introduce new variables in order to be able to map it to an Ising problem. We introduce the variable $a_{i,\alpha}$ which can be either 0 or 1; it is 1 if the salesman goes through city i at step α it is 0 otherwise. So we can write the third condition of the salesman problem as

$$\sum_{\alpha=1}^{N} a_{i,\alpha} = 1 \tag{143}$$

and the second condition as

$$\sum_{i=1}^{N} a_{i,\alpha} = 1 \tag{144}$$

Then the length of the final path can be written as

$$H = \sum_{\alpha=1}^{N} \sum_{1 \le i,j \le N} l_{i,j} a_{i,\alpha} a_{j,\alpha+1} \tag{145}$$

so we are first summing over all steps α and for each of this we considering only the weight $l_{i,j}$ were both $a_{i,\alpha}$ and $a_{j,\alpha+1}$ are 1.

Now we can transform this cost function H into something more similar to what we have for the Ising model

$$H = \frac{1}{4} \sum_{\alpha=1}^{N} \sum_{1 \le i,j \le N} l_{i,j} \sigma_{i,\alpha} \sigma_{j,\alpha+1} + \sum_{\alpha=1}^{N} \sum_{1 \le i,j \le N} l_{i,j} \sigma_{i,\alpha} + C \tag{146}$$

where C is a constant and we used the identity

$$\sigma_{1,\alpha} = 2a_{i,\alpha} - 1 \tag{147}$$

substituting equation (147) in equation (146) gives us back the cost function in equation (145).

Equation (146) is the Hamiltonian of a random Ising model with an inhomogeneous magnetic field.

Other examples on the mapping from NP problem to Ising Hamiltonian can be found in 'Quantum Spin Glasses, Annealing and Computation' [Tanaka et al., 2017].

8 Annealing

8.1 Simulated Annealing

Simulated Annealing or Classical annealing is a computer algorithm that can be used to solve combinatorial optimization problems. This algorithm is not specific to one kind of problem but can be applied to any such type of problem. This algorithm is inspired by statistical physics, and uses the temperature, or thermal fluctuation to find the solution to a given problem. Later we will also see Quantum annealing that makes use of quantum fluctuation to improve simulated annealing.

Simulated annealing has been used in different areas of sciences and has been mathematically proven to work The proof was made by Geman and Geman in one of their paper [Geman and Geman, 1984], and a less rigorous explanation of it can be also found in the before mentioned book by Tanaka et al [Tanaka et al., 2017].

The base concept in simulated annealing is to emulate what happens in the physical annealing as opposed to quenching as explained in section 6. We want to 'cool' the system slowly enough to let it reach the ground state and not get 'stuck' in a relatively high-energy state, a local minimum.

To simulate this physical process with an algorithm the idea by Kirkpatrick described in his paper [Kirkpatrick et al., 1983] is to introduce a temperature term in combination with the optimization problem that we want to solve. This term would have the same role as the physical temperature, hence to assign probabilities to different configurations based on their cost function, as also described in [Stein and Newman, 2013].

One algorithm that can be used to implement this idea is the Metropolis algorithm. The basic idea is to give the algorithm a starting configuration and then a small change to the system is applied,in the case of spin glasses it can be the flip of one spin and the starting configuration can be randomly chosen. Then the energy of the system is calculated and if the new state is lower in energy than the starting one the change is accepted, and this new state becomes the starting configuration for the next iteration. In on the other hand the new state is higher in energy the change is done with a certain probability dependent on the difference in energy between the two state ΔE; this can be taken to be $\exp\{-\frac{\Delta E}{kT}\}$, the Boltzmann factor.

The algorithm then allows it to move from a lower to a higher energy configuration, the reason behind this is that in the case that the system has reached a

local minimum it has the chance to overcome the high potential that separates it from a better minimum, possibly the ground state.

What we would do in quenching is start from a random initial state that represents our system at high temperatures, rapidly go to a low energy state, hence use algorithms such as the Metropolis one with the temperature set to 0. In simulated annealing instead we have to slowly reach the low temperature states, so starting from a high temperature we apply algorithms such as Metropolis for a fixed temperature value, then decrease the temperature and starting from the previously found minimum apply Metropolis again. This is done until we reach 0 temperature; it is important to specify that the decrease in temperature is not linear but the lower the temperature the smaller the decrease per step.

The proof of convergence of the simulated annealing algorithm shows that for a temperature T as a function of the time t the algorithm gives the ground state of the system if

$$T(t) \geq \frac{\Delta}{\ln(t+2)} \tag{148}$$

for some constant Δ depending on the number of sites N (the number of cities if we think at the TSP).

This result is again shown in both the papers we have mentioned at the beginning of the paragraph [Geman and Geman, 1984], [Tanaka et al., 2017].

8.2 Quantum Annealing

In Simulated Annealing we have seen how thermal fluctuations are used in order to escape from a valley or local minimum, and find the global minimum of a system. In its cousin Quantum Annealing a similar method is used, but instead of thermal fluctuation, quantum fluctuation is used, as proposed by Kadowaki and Nishimori [Kadowaki and Nishimori, 1998], as a way to replicate the quantum tunneling effect.

When implementing Quantum annealing we are trying to find the lowest state of a given Hamiltonian. The Hamiltonian takes the form

$$H(t) = H_0 + H_q(t) \tag{149}$$

where the first part of the Hamiltonian H_0 is the Hamiltonian associated to the combinatorial problem we are trying to solve, so it is also associated to an Ising system and has the form

$$H_0 = \sum_{1 \leq i,j \leq N} J_{i,j}\sigma_i \tag{150}$$

where i, j indicate the lattice site and N is the number of sites. So the ground state of H_0 corresponds to the best solution of the optimization problem we are trying to solve.

The time dependent part of the Hamiltonian is the one simulating the quantum fluctuation and can be expressed as

$$H_q(t) = -\Gamma(t) \sum_{i=1}^{N} \sigma_i \tag{151}$$

For large values of $\Gamma(t)$ the quantum effect dominates the system and for $\Gamma(t)$ going to 0 we are back to the simple Ising model Hamiltonian.

What we do practically is to start at large values of $\Gamma(t)$ and start with a trivial ground state of the system; then as $t \to \infty$ we have that $\gamma(t) = 0$ and we obtain the non trivial ground state of our Hamiltonian H_0.

There are different ways of implementing quantum annealing. I am going to give an example of one that makes use of the Schrodinger equation. We start by writing the Schrodinger equation

$$i\frac{d\left|\Psi(t)\right\rangle}{dt} = \hat{H}(t)\left|\Psi(t)\right\rangle \tag{152}$$

where we have taken $\hbar$ to be 1, $|\Psi(t)\rangle$ is the wavefunction of our system at time t and we take $H(t)$ to be the one we defined in equation (149).

We then consider the time evolution operator $\hat{U}(t_1, t_2)$, then the wavefunction at time t is given by

$$\left|\Psi(t)\right\rangle = \hat{U}(t,0)\left|\Psi(0)\right\rangle \tag{153}$$

so by using equation (152) and equation (153) we can write

$$i\frac{d}{dt}\hat{U}(t,0) = \hat{H}(t)\hat{U}(t,0) \tag{154}$$

or

$$\hat{U}(t,0) = 1 - i\int_0^t dt'\hat{H}(t')\hat{U}(t',0) \tag{155}$$

where we used that $\hat{U}(0,0) = 1$. We then can find the solution to equation(155) by rewriting it as

$$\hat{U}(t) = 1 - i\int_0^t dt_1\hat{H}(t_1) \tag{156}$$

$$+ (-i)^2\int_0^t dt_1\int_0^t dt_2\hat{H}(t_1)\hat{H}(t_2) \tag{157}$$

$$+ (-i)^3\int_0^t dt_1\int_0^t dt_2\int_0^t dt_3\hat{H}(t_1)\hat{H}(t_2)\hat{H}(t_3) \tag{158}$$

$$... \tag{159}$$

where we have to notice that $\hat{H}(t_1)$ and $\hat{H}(t_2)$ do not commute for $t_1 \neq t_2$ so we have to use the time ordered operator. Then we get

$$U(t,0) = \sum_0^\infty \frac{1}{n!}(-i)^n\int_0^t dt_1\int_0^t dt_2...\int_0^t dt_n \times \Im\hat{H}(t_1)\hat{H}(t_2)...\hat{H}(t_n) \tag{160}$$

$$= \Im\exp\left\{-i\int_0^t dt'\hat{H}(t')\right\} \tag{161}$$

where Im is the time ordered operator.

In quantum annealing we start by a trivial ground state at $t = 0$, when $Gamma(t)$ is large

$$|\Psi(0)\rangle = |\rightarrow\rightarrow\rightarrow ... \rightarrow\rangle \tag{162}$$

then define $\Gamma(t)$ to be decreasing in t, then calculate $\hat{U}(t, 0)$ and apply it to $|\Psi(0)\rangle$ as in equation(162), for $\Gamma(t) \simeq 0$ so e find the ground state of our optimization problem.

Once we have both simulated and quantum annealing we can compare the two and what we see is that for a fluctuation term, either $T(t)$ or $\Gamma(t)$ that has equation (148) as a lower bound both algorithms reach convergence, so tehy find a solution for the optimization problem, but simulated annealing is faster. On the other hand instead if the lower bound of the fluctuation term is given by

$$\Gamma(t) \geq t^{-\frac{c}{N}} \tag{163}$$

We have that quantum annealing still gives us the solution we are looking for but simulated annealing does not. For simulated annealing we have that the system is 'cooled' down too fast and the algorithm gets stuck in one of the many local minima of the energy landscape.

Graph of these results and their explanation can be found in the appendix

9 Quantum Computers

In this section I will try to give a brief explanation on what quantum computers are and why w e think that they might be so useful in the future. I will then try to give an idea on how much better quantum computers actually are compared to classical one, or if they are better at all. I will then conclude by report prediction on the size of quantum computers in the following year.

The idea of building quantum computers was pushed by the thought that, given the quantum nature of these machines, they would have to be better than classical one in understanding and simulating how molecules interact with each other; and phenomena other interactions that are strongly governed by quantum physics. The first research in quantum computers and quantum algorithms started around the 1980s. One major push in the race of developing these thought to be powerful machines was given by the invention of one of the most famous quantum algorithms. In 1994 Shor developed the so-called Shor's algorithm. This algorithm is used to find the prime factorization of a given number. Thi is something that we were able to do as well on classical computers, but the time it took us to do so is exponential in the number of digits the given number has. This revolutionary algorithm on the other hand is able to solve this problem in polynomial time.

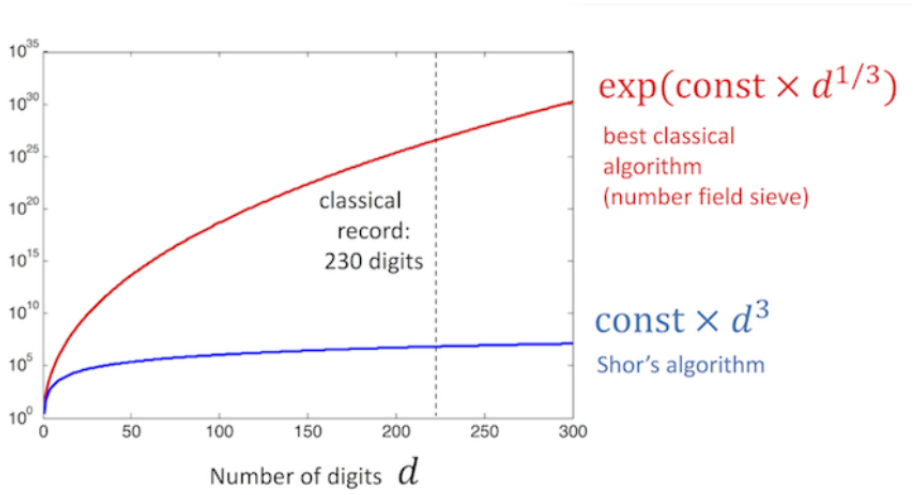

Figure 14: This figure shows the comparison between Shor's algorithm and a classical algorithm for prime decomposition. In the graph we clearly see the exponential behaviour of the classical algorithm versus the polynomial one of Shor's [Quantum,].

So what we see with Shor's algorithm is that a problem we thought to belong to the NP category we have described in chapter 7.1 , is instead part of the N problem. Shor's algorithm works but the problem with it at this moment is that we do not have quantum computers powerful enough to be able to run the algorithms for too large numbers. The implication being able to run such an algorithm would be quite big in today's world. To understand a part of them we have to realize that most of the encryption that is done to safely protect our bank accounts and most of the encrypted data we have today rely on the fact that we are not able to solve the prime factorization of large numbers in a short enough time. We then immediately see that being able to run Shor's algorithm for numbers large enough would put us in a position where our present encryption system would immediately become obsolete. Now this is far from being possible and I will later show some predictions one when we might be able to do such a thing, but just to avoid this to create too many problems scientists are already working on new encryption methods, post Shor's algorithm.

The reason why different algorithms can be developed for quantum computers is that they work in a different way from classical ones. The building blocks of a classical computer are called bits while those of a quantum computer are called qubits. Classical bits encode binary information, each bit can be in one of two states either up or down, or 1 and 0. The qubit, being it a quantum element, is in a superposition of these two states. Once we measure this qubit the output will still be only one of these two states, but we get either one of them with a different probability. Also given the developing stage at which we are with quantum computers there is not a fixed design to build these machines.

This implies that we also do not have a fixed way to build a qubit. A qubit is a quantum object in the sense that it is in a superposition of states. One example could be the spin of an electron, that is not simply either up or down but is in a superposition of these states, one more example is given two atoms that share an electron we can define the '1' state to be the electron being on atom A and the '0' state to be the electron being on the other atom 'B'; so the different ways in which we can physically build a qubit also makes it possible that very different types of quantum computers exists.

No matter the engineering design that is chosen to build the machine qubits share a set of common characteristics such the fact that qubits are entangled with each other, this implies that we cannot look at each separate qubit individual as we might do in a classical computer, but we have to look at the system as a whole. The entanglement of the qubits and the superposition are such that for N qubit we have that the system is in a superposition of 2^N different states. This system is described by the probabilities of each of these 2^N states, hence by 2^N coefficients, so the 'density' of information that we can store in a set of N quantum bits is 2^N versus N in a classical computer.

To give an idea on why we need different algorithms, or better on why we can have different algorithms, for quantum and classical computers lets take entanglement. Two qubits are entangled when the state of one of them depends on the state of the other. If we look at the qubits as if they were spins of an electron, similar reasoning works for different types of qubits as well, we see that given the magnetic field generated from the orientation of one spin the other tends to antiallign to it. The system then is in a lower energy state if these two spins are anti-aligned. Now the four state that can be represented by the two qubit we are considering, those that correspond to the classical state of 00, 01, 10, 11, are the state where the qubits are anti-aligned, no matter in which direction, the state where both of them are up, the states where both of them are down and one more states where they are parallel and both point in the equatorial plane of the system we are considering. Now the main difference here between classical and quantum computers is that if we look at the average of the state where the spins are in opposite directions, as in figure 15, we have that since one spin is in a superposition of up and down and the other one is in the opposite state of the first one, hence it will also be in a superposition of down and up, the average of both spin is 0 in every direction. What we mean is not the state 0, as in the the 0 of the binary code, but we see that the spin is not pointing in any direction, the qubit is on average in none of the possible two states. This is the main difference with a classical computer, there is not a way in classical computers to write a code where the bits are in no state but are anti-aligned with each other, and this is something quantum algorithms can make use of.

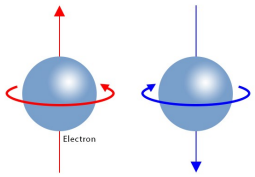

Figure 15: The idea is that the magnetic field lines generated by the orientation of the spin of one electron influence the other one to anti-align. [Learner,].

The operations that we perform on the bits of a computer when we do any sort of computation are physically implemented by logic gates. In a classical computer we can do any operation with only two different types of gate: the not gate that performs operation on one bit to and the NAND gate that performs operation on a pair of bits

Now on quantum computers this second logic gate cannot be used, since this would violate one of the quantum laws that imposes that a quantum system cannot lose information over time and has to be possible to reconstruct the path we have taken to get to a point. A different type of gate has to be used and this is the CNOT gate. A more detailed explanation on how a control not or CNOT gate works can be found in this video [Morello, b].

We have also seen that we are able to process and store more information with the same number of bits than a classical computer but it is hard to 'get it out'. What is usually said is that the power of quantum computers comes from quantum parallelism, that is the possibility to perform operation on a superposition of states. The 'problem' then would be that the final result of the algorithm we have run should be a superposition of states as well, what we need to do is to design an algorithm that instead of giving the system of qubits as a superposition of states it gives them in a more classical version. The system will not actually be classic but we have to find a way such that the probability of one state is much higher than the rest, this state will then be the answer we are looking for. We can better understand this if we look back at the optimization problem and how we mapped their cost function to the Hamiltonian of an Ising system, the output of the quantum computer should give us with a high enough probability a certain configuration, while the initial state is in a superposition of all possible ones, with relatively similar probabilities. One other example is another famous quantum algorithm, called quantum search algorithm, or Grover's Algorithm. This algorithm is designed to find one element among those in an unsorted list. An example is if we have a telephone book where the names are sorted in alphabetical order, we have a number and we want to see who it belongs to. For classical algorithms the only solution we have is to look at each element in the list until we find the one we are looking for. Grover's algorithm starts matching each number with one of the possible states of our qubits system, so it matches it to a combination of 1s and 0s. Then we define a starting wave function that is in a superposition, shown in figure 16 of all these states each with equal probability. What we then do is to perform

a global operation on the wave function that flips the amplitude of the state we are looking for and leaves all the other unchanged. We then apply another operation to the new wavefunction that amplifies the difference between the one we have and the wavefunction of the equal superposition state. The figures 17 and 18 below show a graphic representation of these two passages.

We then repeat these two operations again and we get the the state we are looking for peaks more and more, ,the final wave function then has a much higher probability of being in one state than any other, this is the 'more classical version' we talked about before. The solution of our algorithm is then the states we have the peak for.

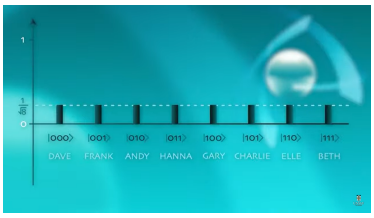

Figure 16: This is an example of Grover's algorithm for $N = 8$ element in the list. This is the representation of the initial wave function where each state has equal probability $\frac{1}{8}$ to be chosen and hence the amplitude of each state is $\frac{1}{\sqrt{8}}$ [Morello, a].

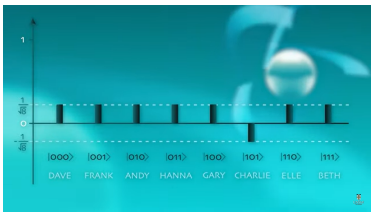

Figure 17: This is the first operation we apply to the wavefunction where we flip only the state we want to find, in this case the state $|101\rangle$, and leave all the other amplitudes untouched [Morello, a].

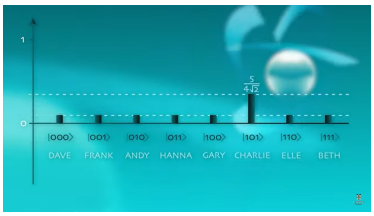

Figure 18: This is the second of the set of two operation that we repeat in each loop and is the one that amplifies the difference between our wavefunction that is the one that is describes in figure 17 and the general wavefunction with all state with equal probability to be chosen, the one in figure 16 [Morello, a].

9.1 D-Wave

What we have described so far is the general structure of a quantum computer, but there are different types of quantum computers, not only in the physical way we build the qubits but also the way the whole computer works. This is the case of D-Wave, a computer that is not part of the category of the so-called Universal quantum computers. There are machines that do not fall into this category and this is the case for D-Wave. D-Wave is a so-called quantum annealer, which is a subclass of adiabatic quantum computers. What this means is that these class of computer follow the adiabatic theorem that states "A physical system remains in its instantaneous eigenstate if a given perturbation is acting on it slowly enough and if there is a gap between the eigenvalue and the rest of the Hamiltonian's spectrum". This is also the same idea that is behind both simulated and quantum annealing. One more difference with D-Wave only is that this as we said is not a universal quantum computer, what this means is that it is built to solve only specific types of problems, it is not a more 'general purpose' computer as the universal one is thought to be.

The main sets of problem D-Wave aims to solve are optimization problems, as the travelling salesman we have seen, and being a quantum annealer it does so mainly by using a combination of quantum and simulated annealing.

9.2 Comparison and Prediction

One thing about quantum computers is that they will not substitute classical ones, but they will be used together to solve specific types of problems, these are problems like quantum simulations, which is the main application quantum computers were first thought for. These would have some of the widest application since quantum simulation is a term that indicates that we would be able to study how different materials interact without having to first create them and then observe them in a laboratory. Some of the possible technological advancements that could follow from working quantum simulations are room temperature superconductors, better drugs and a possible cure for cancer. Other applications are in optimization problems as we have seen in the previ-

ous chapter, application in Artificial Intelligence (AI), financial modelling and others.

Again these are only potential applications since at the moment we have no quantum computer able to implement any of those yet.

Contrary to what many might think at the moment quantum computers seem not to be more powerful than their classical counterparts. The algorithms that we have developed and that we are able to run on quantum computers have revealed themselves better than the one we use for classical computers when the problem that has to be solved is restricted to quite specific cases. The comparison between these two types of machine is usually made using specific types of problems.Quantum computers have proved to be faster for some very specific cases and this was taken to be a result for a more general scenario. As reported by a paper benchmark problem for quantum computers [Mandra et al., 2017] "Either the random spin-glass benchmark problems were too easy on the quasi-planar topology of the D-Wave quantum annealer, or the logical structure of carefully-crafted problems designed to give the annealer an advantages have a trivial structure". What this means is that if we look at the problems that we use to compare quantum and classical computers we see that unless powerful restrictions are applied the best known classical algorithms are still faster than their quantum counterparts as shown in figure **??**.

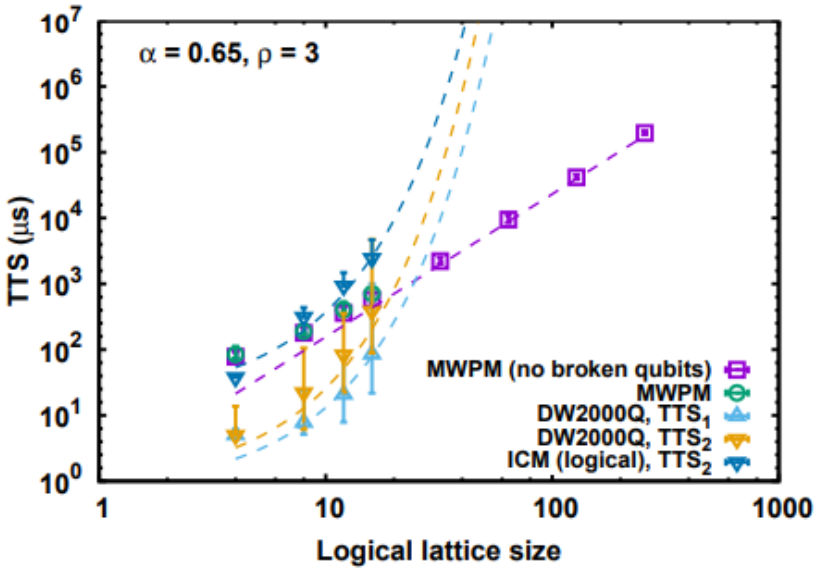

Figure 19: This figure shows the comparison between the best known classical algorithm, in purple, and different trials of quantum algorithms performed on the D-Wave machine [Mandra et al., 2017].

As I said, quantum computers are not going to be used for everyday situations but have some specific kind of problem they are going to focus one. The main reason why we are not able to simulate all sorts of molecule interaction or run Shor's algorithms for any number we want is that today's quantum computers are not powerful enough. The main reasons are noise and the size. The latter presents many engineering challenges linked to how to connect such a huge number of qubits and other sorts of problems, but prediction by some of the main companies that build quantum computers have been made, and are shown in figure 20

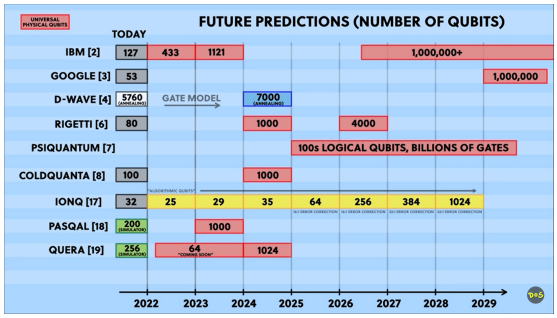

Figure 20: This figure shows some predictions for the next decade on how far some companies should get in building quantum computers. The numbers indicate the number of qubits in each machine and the red boxes indicate physical qubits, the blue one is the prediction for t D-wave, and the yellow ones are logical qubits. The green boxes indicate that those companies at the moment have quantum simulators that can run up to an indicated number of bits, and are planning to build quantum computers [of Science,].

One of the symbolic numbers that we aim to reach in the million qubit computer, that according to the prediction would be able to run error correction code, and also Shor's algorithm, and some relatively complex quantum simulations.

The second problem is what I call noise. What we mean by that is that ideally the qubits we work with are isolated and they do not interact with anything, this is not what happens in reality, this causes the information we stored in them to leak away, this is called decoherence. Tho try to solve this problem quantum error correcting code is thought to be a solution, the basic idea behind this is to use a set of entangled physical qubits to make one physical qubit, now how many we need to use to do so is still not sure but it seems like a number around a hundred or a thousand of physical qubit would do the trick. As I have tried to explain in this last section quantum computers have not reached what we hope is their full potential and the road to that is paved with obstacles, but the possibilities that a fully functioning quantum computer would give us in every area of research are huge.

References

[Baxter, 2016a] Baxter, R. (2016a). *Exactly Solved Models in Statistical Mechanics*. Academic Press.

[Baxter, 2016b] Baxter, R. J. (2016b). *Exactly solved models in statistical mechanics*. Elsevier.

[Castellani and Cavagna, 2005] Castellani, T. and Cavagna, A. (2005). Spin-glass theory for pedestrians. *J. Stat. Mech*, 2005.

[Cipra, 1987] Cipra, B. A. (1987). An introduction to the ising model. *The American Mathematical Monthly*, 94:937–959.

[Edwards and Anderson, 1975] Edwards, S. F. and Anderson, P. W. (1975). Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965.

[Fundation, 2015] Fundation, W. (2015). wikypedia. `https://en.wikipedia.org/wiki/Bethe_lattice`.

[Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

[GmbH, 2002] GmbH, G. (2002). Geogebra. `http://www.geogebra.org`.

[Gualtieri,] Gualtieri, B. tikalon blog by ybob gualtieri. `http://www.tikalon.com/blog/blog.php?article=2017/Horatio_Alger`.

[Guerra, 2004] Guerra, F. (2004). Mathematical aspects of mean field spin glass theory. *arXiv preprint cond-mat/0410435*.

[Kadowaki and Nishimori, 1998] Kadowaki, T. and Nishimori, H. (1998). Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355.

[Kirkpatrick et al., 1983] Kirkpatrick, S., Gelatt Jr, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.

[Learner,] Learner, C. Spin quantum number. `https://www.chemistrylearner.com/spin-quantum-number.html`.

[mandi colloquia,] mandi colloquia, I. Institute online colloquium - dr. steven thomson - the physics of spin glasses - 29th october 2021. `https://www.youtube.com/watch?v=0FTBFCVJOzg&list=PLMtI2BUHjEvGiRNASyN5q4aY7VKeWY2wS&index=5`.

[mandi colloquia, 2021] mandi colloquia, I. (2021). Institute online colloquium - dr. steven thomson - the physics of spin glasses - 29th october 2021. `https://www.youtube.com/watch?v=0FTBFCVJOzg&list=PLMtI2BUHjEvGiRNASyN5q4aY7VKeWY2wS&index=5`.

[Mandra et al., 2017] Mandra, S., Katzgraber, H. G., and Thomas, C. (2017). The pitfalls of planar spin-glass benchmarks: raising the bar for quantum annealers (again). *Quantum Science and Technology*, 2(3):038501.

[Mézard et al., 1987] Mézard, M., Parisi, G., and Virasoro, M. A. (1987). *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company.

[Morello, a] Morello, U.-A. Quantum computing concepts – quantum algorithms. https://www.youtube.com/watch?v=8anVNc0r_8o&list=PLHSIfioizVW2uC27IFkHlSc-NgvZjBliZ&index=8.

[Morello, b] Morello, U.-A. Quantum computing concepts – quantum logic. https://www.youtube.com/watch?v=YTNug9tQOzU&list=PLHSIfioizVW2uC27IFkHlSc-NgvZjBliZ&index=7.

[Nishimori, 2001] Nishimori, H. (2001). *Statistical physics of spin glasses and information processing: an introduction*. Number 111. Clarendon Press.

[of Science,] of Science, D. Who has the best quantum computer? https://www.youtube.com/watch?v=gcbMKt079l8.

[Peierls, 1936] Peierls, R. (1936). On ising's model of ferromagnetism. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 32, pages 477–481. Cambridge University Press.

[Quantum,] Quantum, I. Shor's algorithm. https://quantum-computing.ibm.com/composer/docs/iqx/guide/shors-algorithm.

[Stein and Newman, 2013] Stein, D. L. and Newman, C. M. (2013). *Spin glasses and complexity*, volume 4. Princeton University Press.

[Sürmeli,] Sürmeli, C. P vs. np vs. np-complete vs. np-hard. https://www.cansurmeli.com/posts/p-vs-np-vs-np-complete-vs-np-hard/.

[Tanaka et al., 2017] Tanaka, S., Tamura, R., and Chakrabarti, B. K. (2017). *Quantum spin glasses, annealing and computation*. Cambridge University Press.

[Wald, 2017] Wald, S. (2017). Thermalisation and relaxation of quantum systems.

[Wikipedia contributors, 2022] Wikipedia contributors (2022). Descartes' rule of signs — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Descartes%27_rule_of_signs&oldid=1096620783. [Online; accessed 14-July-2022].

10 Appendix

Following the definition of dimensionality used in (76) we were able to show that the Bethe lattice is an infinite dimensional lattice. Here I show that this definition holds also for more common 2 and three dimensional lattices

10.1 2D square lattice

The easiest non one dimensional lattice we can think of is the two dimensional square lattice. To use the formula (76) we have to know how many lattice site we can reach within n steps.

With 0 steps we can only reach the site from which we start, with one step we reach the 4 sites directly connected to it, and we see that with each next step we can reach all the sites within a square of side $n\sqrt{2}$, tilted by 90 degrees, see figure 7. So the number of new sites we can reach at the nth step is 4n hence the number c_m in (75) is

$$c_m = 1 + 4 + 8 + 12+ \tag{164}$$

$$= 1 + \sum_{n=1} 4n \tag{165}$$

$$= 1 + 2n(n+1) \tag{166}$$

Using (76)

$$d = \lim_{n\to\infty} \frac{\ln(1 + 2n(n+1))}{ln(n)} \tag{167}$$

$$\sim \lim_{n\to\infty} \frac{\ln(2n(n+1))}{ln(n)} \tag{168}$$

$$= \lim_{n\to\infty} \frac{\ln(2n) + \ln(n+1)}{ln(n)} \tag{169}$$

$$= 2 \tag{170}$$

10.2 3D cubic lattice

We can do the same as we did before and get the number c_m for a 3D cubic lattice. There are many ways to approach this question, one could use a similar reasoning as the previous one with a cube instead of a square, but the following approach is maybe less rigorous. We think of a £D cubic lattice as many layers of a 2D square lattice. We consider the number of lattice sites that can be reached within exactly n steps, from a fixed point that we can call O. We start looking at the horizontal plane in which we have O, here we have the exact number of new sites that we had after n steps on a 2D lattice. When we have to consider another horizontal plane, we have to take into account that to reach' that plane we spend in steps the distance of O to that plane, we should call this distance t. The number of new neighbours for each plane is the same as a 2D

lattice at the step $n-t$. One last step before being able to determine c_m is to notice that, beside the 'first' horizontal plane in which there is O, we have two planes at distance t, one 'above' and one below' our 'first plane.

New we can write m_i in (75) as

$$m_i(3D) = 2 \cdot \left[\sum_{j=0}^{j=i} m_j(2D) \right] - m_i(2D) \tag{171}$$

$$= 2c_i(2D) - m_i(2D) \tag{172}$$

$$= 2 + 4i(i+1) - 4i \tag{173}$$

Where $(3D)$ or $(2D)$ indicates which lattice we are referring to, either the square or the cubic one. The formula is valid for $i > 0$, for $i = 0$ $m_i = 1$.

Now we calculate c_m

$$c_m = \sum_{i=0}^{n} m_i \tag{174}$$

$$= \left[\sum_{i=1}^{n} 4i(i+1) \right] + (2(n+1)) - (2n(n+1)+1) \tag{175}$$

$$= 4 \left[\sum_{i=1}^{n} i^2 + 1 \right] + (2(n+1)) - (2n(n+1)+1) \tag{176}$$

$$= 4 \left[\sum_{i=1}^{n} i^2 + \sum_{i=1}^{n} i \right] + (2(n+1)) - (2n(n+1)+1) \tag{177}$$

$$= 4 \left[\frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)}{2} \right] + (2(n+1)) - (2n(n+1)+1) \tag{178}$$

$$= 2\frac{n(n+1)(2n+1)}{3} + 2n(n+1) + (2(n+1)) - (2n(n+1)+1) \tag{179}$$

$$= 2\frac{n(n+1)(2n+1)}{3} + 2n + 1 \tag{180}$$

Now we can calculate the dimensions of this lattice

$$d = \lim_{n\to\infty} \frac{\ln\left(\frac{2}{3}n(n+1)(2n+1) + 2n + 1\right)}{\ln(n)} \tag{181}$$

$$\sim \lim_{n\to\infty} \frac{\ln\left(\frac{2}{3}n(n+1)(2n+1)\right)}{\ln(n)} \tag{182}$$

$$= \lim_{n\to\infty} \frac{\ln\left(\frac{2}{3}\right) + \ln(n) + \ln(n+1) + \ln(2n+1)}{\ln(n)} \tag{183}$$

$$= 3 \tag{184}$$

10.3 Graphs of quantum and simulated annealing

In this subsection we show the result of the comparison of simulated and quantum annealing. All the graphs have been taken from a book by Tanaka et al

cited in the references [Tanaka et al., 2017], which took them themselves from the article of Kadowaki and Nishimori [Kadowaki and Nishimori, 1998].

What the following graphs show is the comparison of quantum annealing and simulated annealing as we change the time we take to run the algorithms. As we know we have that the convergence theorem for both algorithms holds only if the 'cooling' is done slowly enough, the lower bounds are indicated in equations (148) and (163).

We indicate by P_{SA} and P_{QA} the time depend probabilities of the system being in the ground state of simulated and quantum annealing respectively, note that these comparison have been performer on system for which we were able to calculate by other methods the ground state so that we know what the actual answer is.

Two other parameters that are used for the comparison are P^{st}_{SA} and P^{st}_{QA} that are the prediction of what the probabilities should be, their meaning is better explained in the above cited papers. What is of interest for us is the difference between P_{SA} and P^{st}_{SA} or the quantum analog of this quantity. If indeed these show different curves as the time goes to large values it means that the algorithm has not reached the ground state but it is stuck in one of the many local minimums of the system.

The following graphs show how the two algorithms behave under three different 'cooling', so the system is cooled faster in each successive graph according to the following equations

$$\frac{\Gamma(t)}{J} = \frac{T(t)}{J} = \frac{3}{\ln(t+1)} \tag{185}$$

$$\frac{\Gamma(t)}{J} = \frac{T(t)}{J} = \frac{3}{\sqrt{t}} \tag{186}$$

$$\frac{\Gamma(t)}{J} = \frac{T(t)}{J} = \frac{3}{t} \tag{187}$$

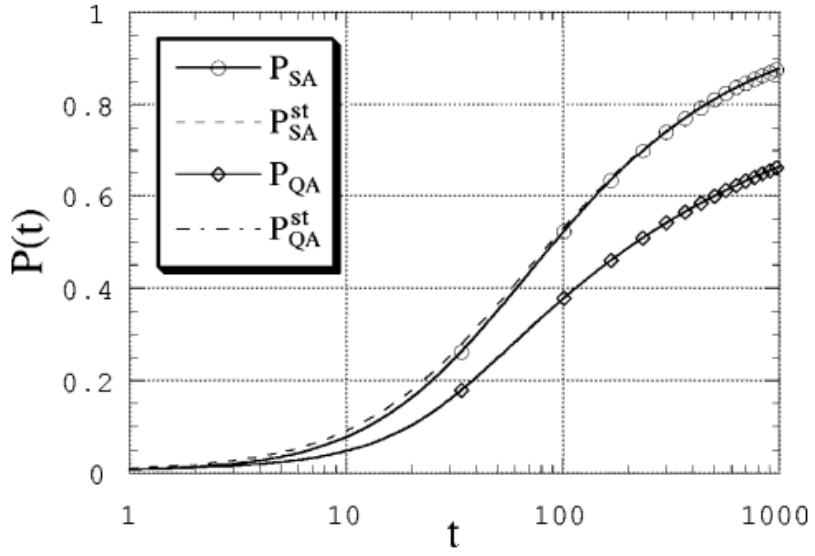

Figure 21:

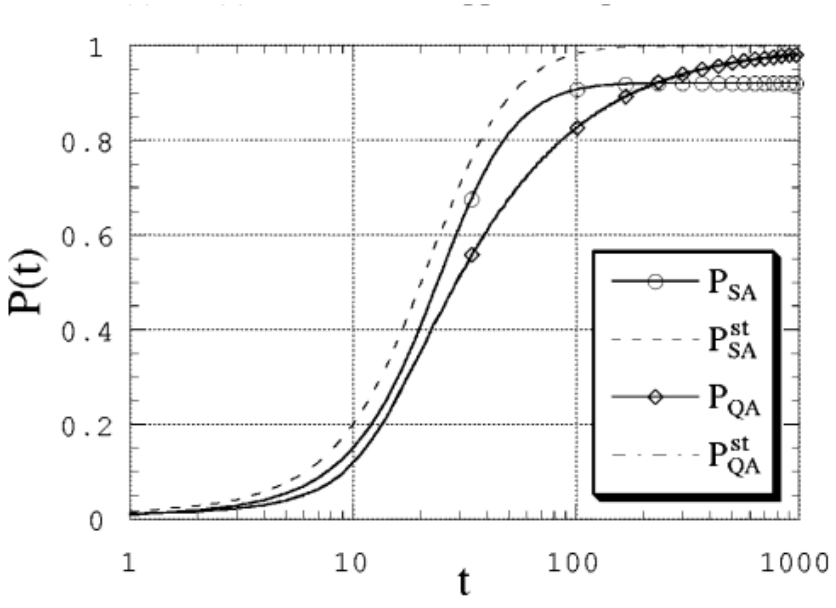

Figure 22:

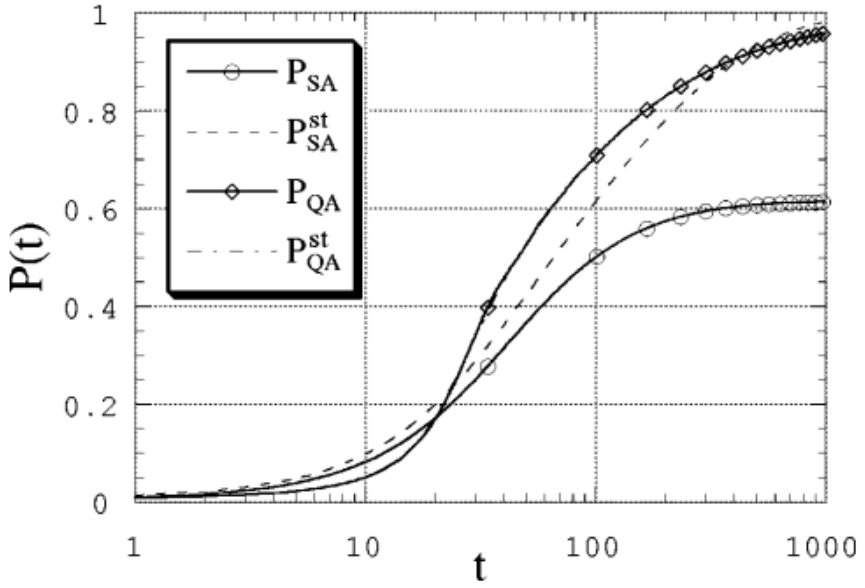

Figure 23:

What we see in the first graph, cooled according to equation (**??**) is that both algorithm reach the ground state since we sew overlapping between P_i and P_i^{st}, but we see that simulated annealing reaches a better result since gives us a higher probability of finding the ground state. Equation (**??**) is within the limit allowed by the convergence algorithm of simulated annealing, but the next two cases are not.

In the second graph we see as for the third one that quantum annealing reaches a relatively high probability of giving us the ground state and also simulated annealing does not converge to the ground state