Statistical Physics
SISSA – International School for Advanced Studies

The Transverse-Field Ising Spin Glass Model on the Bethe Lattice

with an Application to Adiabatic Quantum Computing

Gianni Mossi

Thesis submitted for the degree of
Doctor of Philosophy

4 October 2017

Advisor: Antonello Scardicchio

Abstract

In this Ph.D. thesis we examine the Adiabatic Quantum Algorithm from the point of view of statistical and condensed matter physics. We do this by studying the transverse-field Ising spin glass model defined on the Bethe lattice, which is of independent interest to both the physics community and the quantum computation community. Using quantum Monte Carlo methods, we perform an extensive study of the the ground-state properties of the model, including the Rényi entanglement entropy, quantum Fisher information, Edwards–Anderson parameter, correlation functions.

Through the finite-size scaling of these quantities we find multiple independent and coinciding estimates for the critical point of the glassy phase transition at zero temperature, which completes the phase diagram of the model as was previously known in the literature. We find volumetric bipartite and finite multipartite entanglement for all values of the transverse field considered, both in the paramagnetic and in the glassy phase, and at criticality. We discuss their implication with respect to quantum computing.

By writing a perturbative expansion in the large transverse field regime we develop a mean-field quasiparticle theory that explains the numerical data. The emerging picture is that of degenerate bands of localized quasiparticle excitations on top of a vacuum. The perturbative energy corrections to these bands are given by pair creation/annihilation and hopping processes of the quasiparticles on the Bethe lattice. The transition to the glassy phase is explained as a crossing of the energy level of the vacuum with one of the bands, so that creation of quasiparticles becomes energetically favoured.

We also study the localization properties of the model by employing the forward scattering approximation of the locator expansion, which we compute using a numerical transfer matrix technique. We obtain a lower bound for the mobility edge of the system. We find a localized region inside of the glassy phase and we discuss the consequences of its presence for the Adiabatic Quantum Algorithm.

Preamble

Motivations

Adiabatic quantum computing was developed in 2001 as an alternative to the circuit model of quantum computation. It is a topic where many diverse areas of physics meet. Quantum phase transitions, the physics of disordered and glassy systems, out-of-equilibrium dynamics, entanglement and quantum information theory all have a role to play. Moreover, it is an interdisciplinary field related to scientific fields such as mathematics and theoretical computer science. From the perspective of the theoretical physicist, this allows us to test the methods of quantum physics into uncharted territories for which they were not designed, often resulting in a new and better understanding of the theory.

However, adiabatic computing is not only of theoretical interest. Quantum computing crossed from academia and into the industry when the D-Wave company developed the first commercial quantum adiabatic computer in 2011. Stimulated by the possibilities offered by real-life quantum computational devices, a lot of effort went into trying to understand the physics and the working of the D-Wave machine, as well as the theoretical reasons why adding quantum mechanics to analog computational devices could improve their performance.

Finally, from a computer science perspective, adiabatic quantum computing provides us with a different and original way of looking at quantum algorithms. After about two decades of research, the number of quantum algorithms we have is still relatively small. The consensus in the scientific community is that quantum algorithms are hard to devise. One of the advantages of the Adiabatic Quantum Algorithm is that it allows people to use physical intuitions in order to get insights into quantum computation.

It is with these goals in mind that the author of this thesis, whose scientific background was mostly mathematics and computer science related, decided to apply to a Ph.D. program in physics in 2013.

Thesis Outline

This thesis is organized in four chapters. The first chapter introduces the theoretical background from which the thesis starts. The adiabatic quantum algorithm is presented, as well as the physical mechanisms underpinning its workings and the physical models that have been proposed to better understand it. We also review the entanglement measures that will be relevant to us in later chapters.

The second chapter introduces the classical Ising spin glass on the Bethe lattice, its solution through the cavity method and its connection with classical computational complexity.

In the third chapter the Ising spin glass is quantized by the introduction of a transverse field. The zero-temperature thermodynamic properties of the moden are studied numerically and the critical value of the glassy transition is extracted from them. In order to understand the critical point, a quasiparticle picture is developed from the perturbative expansion in the large transverse field limit.

The fourth chapter briefly introduces the Anderson model and the connections between localization and the Adiabatic Quantum Computing. The localization properties of the transverse-field Ising spin glass model are studied using the forward scattering approximation. Conclusions follow.

An extensive appendix collects the details on the numerical methods that have been employed throughout the thesis, but were not included in the main text for a better readability.

The original research of the thesis is in Chapter 3 [1] and Chapter 4 [2].

Acknowledgements

I would like to thank my scientific advisor Professor Antonello Scardicchio of ICTP, for all the help and support he's given me during these years of grad school, as well as my co-authors Dr. Sebastiano Pilati and Tommaso Parolini, and all teaching and research staff in SISSA and ICTP.

I thank my family for their support: Sergio, Marco and Darma. I thank my grandmother Luciana who passed away during this time I spent in Trieste and abroad.

I thank all my friends and collegues here in Trieste, in particular Cristiano de Nobili, Ivan Girardi, Giulia Rossi, Valeria Mantese, Amelia Monini, Sara Bertola. Special mention for the people who have shared an office with me: Francesca Pietracaprina, Leda Bucciantini, Urna Basu, Alessio Lerose, Marlon Brenes, Giorgio Zicari, Saranyo "Nino" Moitra and last but not least Silvia Pappalardi, to whom I finally relinquish the seniorship of office 262 in ICTP.

Contents

Chapter 1

Introduction and Theoretical Background

In this chapter we present the background material from which this thesis starts. In Section 1.1 we introduce the historical perspective on quantum computing that led to the developement of the quantum adiabatic algorithm and provide some motivations for the theoretical choices and the line of reasearch we take in this thesis. In Section 1.2 we describe in detail the quantum adiabatic algorithm and the combinatorial optimization problems it is designed to solve. In Section 1.3 we consider the physical interpretations that have been proposed in order to understand the physics behind its workings. In Section 1.4 we review the connections between entanglement and quantum computation that will be of interest in later chapters.

1.1 Quantum Computation

The birth of quantum computation is usually traced back to Feynman's 1982 paper [3] where the concept of a computer that is quantum mechanical in nature was proposed in order to efficiently simulate the real-time dynamics of quantum mechanical systems. The main intuition behind this idea is that the Hilbert space of a many-body system has a dimension that grows exponentially in the number of particles, therefore a state of the system takes an exponential amount of resources just to be stored in the memory of a classical, digital computer. Feynman noted that by using a memory register made of *qubits* (*i.e.* two-dimensional quantum systems) instead of bits one could store a many-body quantum state at linear – rather than exponential – cost. Hamiltonians would then be engineered so that the state encoded in the register would evolve according to the dynamics of the system one is interested in simulating.

In the following decades scientists pondered whether this hypothetical quantum computer could be useful for more than just quantum simulations, and the standard "quantum circuit" architecture for quantum computers was developed in order to tackle more general computational problems. A quantum circuit applies a sequence of one or two-qubits unitary gates to a working register of qubits, in a similar manner to how transistors in electronic circuits apply logical operations to a register of bits. With this analogy in mind, scientists studied

whether adding quantum mechanics would allow a computer to perform computational tasks that are beyond the power of a classical computer. While the strict answer to this question turned out to be "no", as the set of quantum-computable functions coincides with the set of Turing-computable functions, a more interesting results was found within the field of computational complexity, a sub-branch of theoretical computer science that studies the minimal resources needed to perform a specific computational task. Currently, a quantum computer stands as the most likely counterexample to the *extended Church-Turing thesis*, a hypothesis according to which all "reasonable" computational model can be efficiently simulated by a standard Turing machine. In addition to the simulation of quantum physics proposed by Feynman, in 1994 Peter Shor discovered a quantum algorithm capable of efficiently factoring integers, a problem that is not known to have efficient classical algorithms and whose hardness is a necessary condition for the security of commonly-used cryptosystems such as RSA. This result generated a surge of interest in quantum computing and led to the discovery of a number of quantum algorithms.

In 1998, Kadowaki and Nishimori [4] proposed an approach to quantum computing different from the circuit model, an approach that is closer to the spirit of condensed matter physics. Their idea is to use quantum fluctuations to efficiently explore the energy landscape of a classical physical system, in a process – now called quantum annealing (QA) – that is similar in spirit to thermal annealing, *i.e.* the gradual cooling of a system in order to reach its ground state. Starting with a wavefuction spread over all the states in a classical potential, strong quantum tunnelling processes would be turned on by fine-tunining a paramater in the Hamiltonian. When tunnelling rates are very high, the system is able to efficiently explore all of its Hilbert space. The tunnelling processes are then gradually attenuated to zero and the system ends up trapped in a classical low-energy state (ideally its ground state) if this is done slowly enough. In 2000 Farhi *et al.* [5] independently devised a generalization of quantum annealing called the adiabatic quantum algorithm (AQA). Despite its name, it is not a complete algorithm but rather a metaheuristic[1]. The AQA works by a method known as adiabatic state preparation, where the ground state of a target Hamiltonian H_P is obtained by the real-time dynamics generated by a quasi-static time-dependent Hamiltonian $H(t)$ that connects an initial Hamiltonian H_0 to H_P. If $H(t)$ changes slowly enough, then its dynamics will evolve the ground state of H_0 into the ground state of H_P, which in the case of the AQA will be engineered so as to encode the solution to a given computational task one wants to perform. The details of this procedure are described in Section 1.2. Adiabatic quantum computing was proved to be equivalent (up to a polynomial overhead) to the more standard circuit model of quantum computation [6]. Interestingly, the proof relies on methods similar to the ones used in the proof of the Cook-Levin Theorem [7], a well-known theorem in classical computational complexity. This result can be used to map any quantum algorithm written in the quantum circuit model to a specific implementation of the adiabatic algorithm. This however remains only of academic interest as the time-dependent Hamiltonian produced by this mapping contains complicated terms whose physical implementation is well beyond the possibility of current technologies.

[1] from μετά "beyond", and εὑρίσκειν "to discover". As the prefix *meta* is commonly used to describe a higher level of abstraction than the one used in the object language, we can loosely define a "metaheuristic" as "a template for the formulation (or discovery) of new algorithms."

One of the attractive features of the AQA is its wide applicability to different optimization problems. The flip side of this coin is that, being essentially a computational model and not a complete algorithm devised for the solution of a specific problem, it is not at all obvious on which problems the AQA performs best. Since its inception, scientists have tried to find a problem at which the AQA could perform demostrably better that all known classical algorithms. The original paper by Farhi *et al.* proposed to use the AQA to solve NP-hard problems and the early numerical results showed a runtime that scaled polynomially with the instance size [5, 8]. It became quickly evident that this was a pre-asymptotic behaviour and the study of larger sizes showed an exponential relation between the running time and the size of the instance [9, 10, 11]. At the same time, complexity theory results were proved that made such a possibility quite remote.

In order to benchmark the average-case computational power of the algorithm, and following the influential tunnelling interpretation of the quantum annealing process, people started to compare the respective preformances of quantum annealing and thermal annealing over different optimization problems, hoping to find some example of a quantum speedup. The usual theoretical arguments for a comparison of the two processes is reviewed in Section 1.2. Most of these comparison were conducted by simulation. In particular, in [12] Martonak *et al.* used Quantum Monte Carlo methods to simulate runs of the quantum annealing protocol, and showed an asymptotic speedup of simulated quantum annealing over simulated thermal annealing. The problem considered in the paper was a transverse-field spin glass with interaction defined on a square lattice. This effect was later found to be by no means universal, as subsequent results [13] showed for example that on k-SAT simulated annealing seems to perform better than simulated quantum annealing. In [14], Heim *et al.* argued that the observed speedup was a spurious byproduct of the numerical methods used (specifically, the discretization of the imaginary time). Following this result, the statistical analysis of the performance of the AQA over random instances of optimization problems seems to have lost some of its momentum.

Nowadays most of the quantum computing community believes that quantum computers cannot solve NP-hard problems in polynomial time, in the worst case. Finding a *natural, useful* problem where the AQA is provably exponentially better than all known classical algorithms is still an open problem in the field. Rather than try to find a collection of computer-simulation techniques – *e.g.* the many Monte Carlo variants – that reproduces the input-output statistics of a real-life quantum annealer like the D-Wave without actually following its dynamics (which essentially means treating the annealer as a black box), current efforts are exploring the possible physical mechanisms that might give quantum annealing an advantage over classical computation. For this goal it seems that the out-of-equilibrium, real-time evolution of a quantum annealing process needs to be studied in greater detail.

1.1.1 The D-Wave Machine and our theoretical model

The focus of this thesis is the Ising spin glass model in a transverse field with nearest-neighbours interactions defined on the Bethe lattice, an infinite regular

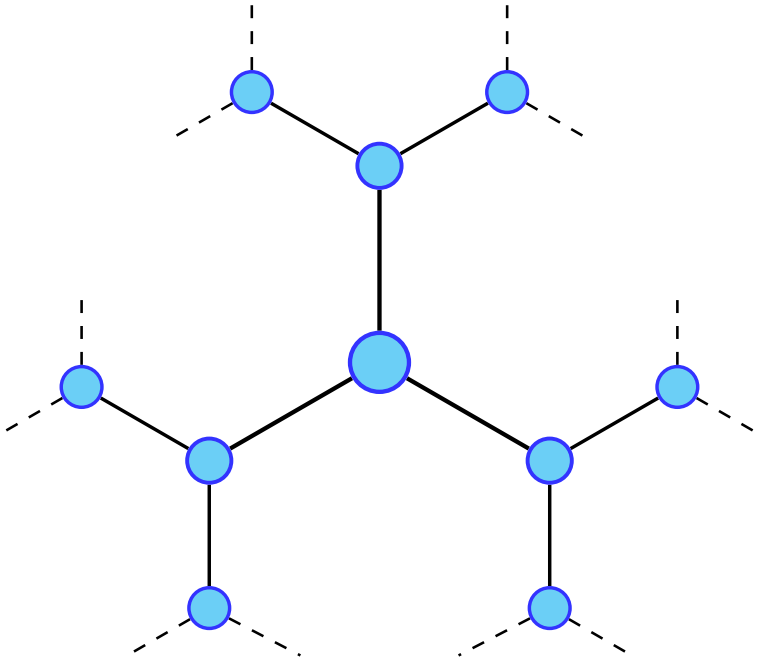


Figure 1.1: The Bethe lattice of branching number K is the unique connected, $(K+1)$-regular tree on an infinite number of nodes, depicted above for $K = 2$. Starting from any point in the graph, the r-th "shell" around the point has $(K+1)K^{r-1}$ elements, which means that in the truncated tree only a fraction $\sim 1/K$ of the spins belongs in the bulk.

tree shown in Fig. 1.1. Its Hamiltonian is the following.

$$H = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i^z\sigma_j^z - \Gamma\sum_i \sigma_i^x. \tag{1.1}$$

The reasons we chose to study this model are related to the current architecture of real-life quantum annealers. As of 2017, adiabatic quantum computation has become the preminent model of pratical quantum computation, and this is almost singlehandedly due to a Canadian company called D-Wave, that in 2011 produced the first quantum annealer using superconducting flux qubits coupled together. The chip of the D-Wave machine can be envisioned as a flat surface tiled with square unit cells. Inside of a unit cell are embedded long and thin superconducting loops, arranged in a crisscrossing array shown in Fig. 1.2. The direction of the current inside of each circuit is the binary degree of freedom that physically represents one qubit, with superpositions of currents representing the generic qubit state $\alpha|0\rangle + \beta|1\rangle$. Different loops can be coupled together only if they intersect at some point. This produces an effective term $-J_{ij}\sigma_i^z\sigma_j^z$ with a tunable intensity J_{ij} controlled by the strength of the coupling between the loops. The same physical considerations taken at the boundary of each cell

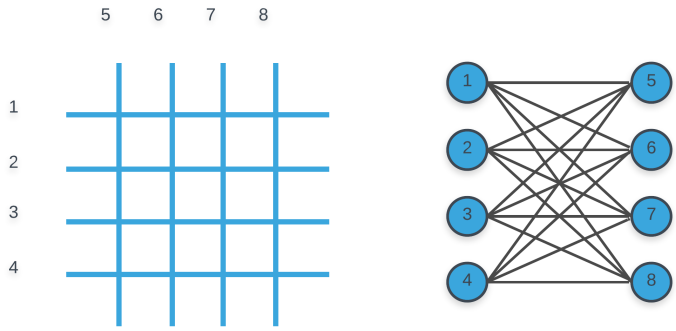


Figure 1.2: schematic representation of the D-Wave unit cell. (Left) Horizontal and vertical lines represent long and narrow superconducting loops, which are associated to effective spins. Only intersecting loops can be coupled together. (Right) Graph representation of the effective spins in the unit cell: nodes represent spins and links between nodes represent interactions between the spins.

show *e.g.* that a horizontal superconducting loops can be coupled only to the corresponding horizontal loops in the nearest cells to its left and to its right, and similarly for the vertical loops. This defines the Chimera interaction graph between spins, shown in Fig. 1.3. A magnetic field in a direction orthogonal to the plane of the chip provides the transverse-field term $-\Gamma \sum_i \sigma_i^x$, obtaining the full Hamiltonian above.

The couplings between the spins can be fine-tuned so as to define a spin glass Hamiltonian. In a typical run of the machine, a strong transverse field is turned on while the spin-spin interactions are absent and the system is cooled down to ≈ 80 mK. Then the transverse field is slowly decreased to zero while the interactions are slowly turned on. While the behaviour of the D-Wave machine is not completely clear and its status as a quantum computing device is somewhat controversial (see *e.g.* [15, 16, 17, 18]), its impressive working register of 2000 qubits allows for the use in practical machine learning problems [19, 20, 21], after these are translated into instance of the native spin glass problem that the D-Wave machine is able to solve.

The model we study, however, is *not* defined on the Chimera graph, for the following reasons. Even though the Bethe lattice and the Chimera graph are quite different, we believe the model of Eq. (1.1) is a good compromise between a system that is of independent interest to the physics community, has a well-defined place in the scientific literature and is amenable to various numerical and analytic techniques on one hand, and a system that faithfully reflects the features of real-life quantum annealers such as the D-Wave machine, and where one can derive proof-of-concept results that are ultimately applicable to these practical devices, on the other.

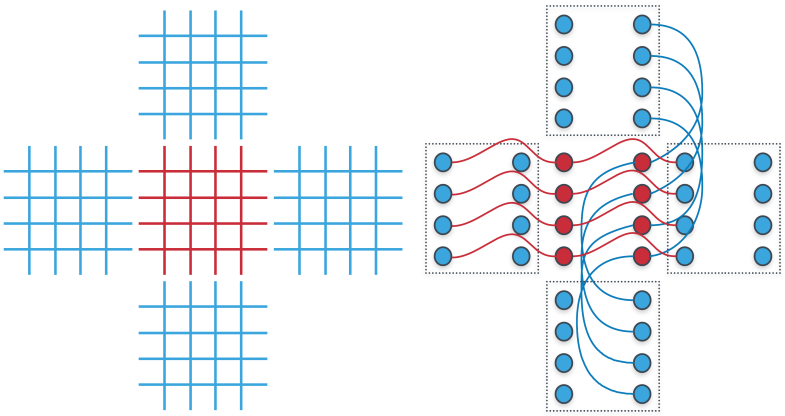


Figure 1.3: (Left) physical array of four unit cells (blue lines) sorrounding a given unit cell (red lines) in a D-Wave chip. In addition to the couplings within the red cell itself (see Fig. 1.2), a red horizontal loop is coupled to the two horizontal loops directly to its left and right. A red vertical loop is coupled to the two vertical loops directly above and below. (Right) Graph representation of the effective spin-spin interactions between between different unit cells. Red links signify interactions across neighbouring unit cells in the horizontal direction, while blue links signify interactions across neighbouring unit cells in the vertical direction.

1.2 Adiabatic Quantum Computing

In this Section we describe in detail the adiabatic quantum algorithm and the computational problems it was designed to solve.

1.2.1 Combinatorial Optimization

The Adiabatic Algorithm is a metaheuristic for the solution of *combinatorial optimization problems* (COPs). These problems are defined by a space of binary strings $\mathbb{B}_N = \{0,1\}^N$ of length N and a cost function $f : \mathbb{B}_N \to \mathbb{R}$ that assign a "cost" $f(x)$ to each string $x \in \mathbb{B}_N$. In order to solve the problem we are asked to find a string x^* that minimizes[2] $f(x)$ over the variational space $\mathbb{B}_N$:

$$f(x^*) = \min \Big\{ f(x) \mid x \in \mathbb{B}_N \Big\}$$

In most practical cases (and in all the cases we will consider) the function $f(x)$ is given as a sum of polynomially-many (in the number of bits N) functions f_i that depend only on a constant number k of variables $x_{j_1^{(i)}}, \ldots, x_{j_k^{(i)}}$:

$$f(x_1, \ldots, x_N) = \sum_{i=1}^{m} f_i(x_{j_1^{(i)}}, \ldots, x_{j_k^{(i)}}).$$

[2] or maximizes, depending on the specific problem. Since maximizing $f(x)$ is equivalent with minimizing $-f(x)$ we assume without loss of generality that we are always dealing with minimization problems.

The integer k is the "locality" of the function f_i and we call the function f an instance of a "k-local" constraint satisfaction problem. When considering collections of COPs with different sizes N, we make the additional technical assumption that the costs $f_i(x)$ do not grow with N. Local COPs are ubiquitous in mathematics, physics, computer science, engineering and many other fields.

Combinatorial optimization problems are related to *constraint satisfaction problems*, which are COPS where the the functions f_i are called *constraints* and take values in $\{0,1\}$. A string x such that $f_i(x) = 0$ is said to *satisfy* the constraint f_i, while $f_i(x) = 1$ means that the constraint is *violated*. COPs are also related to *decision problems*, which are computational problems that admit a yes/no answer. There is a canonical way to formulate any COP as a decision problem: given a COP instance f and a real number $a \in \mathbb{R}$, one is asked to decide whether there is a string $x \in \{0,1\}^N$ such that $f(x) < a$. As an example, we mention the decision problem Exact-Cover as this is a problem we will encounter later.

The Exact-Cover problem can be formulated as follows. Given

- a finite set $S = \{s_1, \ldots, s_N\}$, and
- a collection $\mathcal{A} = \{A_1, \ldots, A_M\}$ of subsets of S, where $M = \mathsf{poly}(N)$,

one is asked to decide whether there is a way of choosing subsets $A_1, \ldots, A_k \in \mathcal{A}$ so as to form a partition of S, meaning that each element $s_i \in S$ belongs to exactly one subset in the chosen subcollection $\mathcal{A}' \equiv \{A'_1, \ldots, A'_k\} \subseteq \mathcal{A}$. This problem is often described in the *inverse representation*, where a candidate solution for the problem (*i.e.* a subcollection $\mathcal{A}' = \{A_{i_1}, \ldots, A_{i_k}\} \subseteq \mathcal{A}$) is described by a binary string. This is done by defining, for each set $A_j \in \mathcal{A}$, a Boolean variable X_j. Then the candidate solution $\mathcal{A}' = \{A_{i_1}, \ldots, A_{i_k}\}$ is represented by the M-length binary string $X = (X_1, \ldots, X_M) \in \{0,1\}^M$ where the bits $X_{i_1}, \ldots, X_{i_k}$ are equal to one, while all the other bits are equal to zero (*i.e.* the value of the variable X_j tells if the set A_j belongs to the given subcollection $\mathcal{A}'$ or not). In this way one defines a bijection between $\{0,1\}^M$ and $\mathcal{A}$, so that a binary string $X \in \{0,1\}^M$ can be univocally associated to a subcollection $\mathcal{A}'_X \subseteq \mathcal{A}$ and *vice versa*.

One can then rephrase the statement "$\mathcal{A}'$ is a partition of the set S" as a formal property of binary strings as follows. Let s_i be an element of S, and let $A_1^{(i)}, \ldots, A_p^{(i)}$ be the subsets in $\mathcal{A}$ that contain s_i. Note that a necessary condition for a subcollection $\mathcal{A}'$ to be a partition of S is that $\mathcal{A}'$ must contain exactly one of these subsets. In the binary string language, this means that the string X must have exactly one of the bits $X_1^{(i)}, \ldots, X_p^{(i)}$ equal to one. We can enforce this by defining a constraint $f_i(X_1^{(i)}, \ldots, X_p^{(i)})$ that is satisfied if and only if exactly one of the variables $X_1^{(i)}, \ldots, X_p^{(i)}$ has a value of one, and is violated otherwise. This gives us N constraints, one for each element $s_i \in S$. The collection $\{f_i\}_{i=1}^N$ of these constraints completely captures the property of being a solution to the Exact-Cover problem, in the sense that for all binary strings $X \in \{0,1\}^M$ it holds that (under the $X \leftrightarrow \mathcal{A}'_X$ association defined above)

$$\mathcal{A}'_X \text{ is a partition of } S \iff f_i(X) = 0 \text{ for all } 1 \leq i \leq N.$$

The restricted problem 3-Exact-Cover requires in addition that each element s_i must be contained in exactly three sets of $\mathcal{A}$, *i.e.* $m = 3$ for each clause

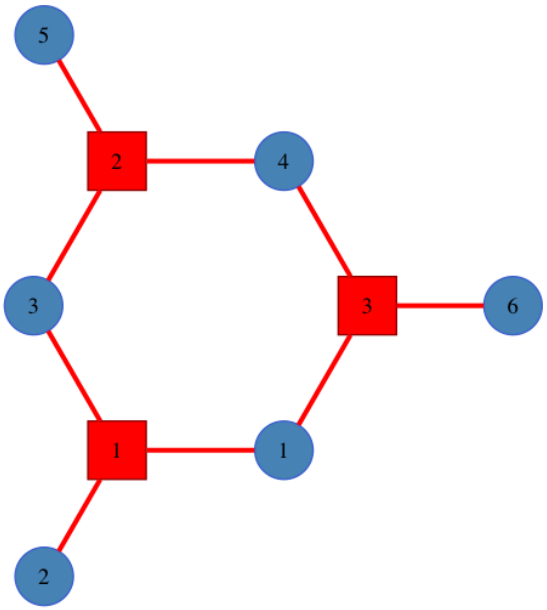


Figure 1.4: example of the factor graph of an 3-EXACT-COVER instance over the Boolean variables $X_1, \ldots, X_6$ (shown here as blue round vertices) with factors $f_1(X_1, X_2, X_3)$, $f_2(X_3, X_4, X_5)$ and $f_3(X_1, X_4, X_6)$ (shown as red squares). Factor functions for 3-EXACT-COVER are of the form $f(X, Y, Z) = (X+Y+Z-1)^2$.

f_i.

We can cast 3-EXACT-COVER as an optimization problem by defining the cost function

$$f_i(X_1, X_2, X_3) \equiv (X_1 + X_2 + X_3 - 1)^2.$$

for each constraint in the system. The total cost function is then $f = \sum_i f_i$. Solutions are binary strings $x \in \{0, 1\}^N$ such that $f(x) = 0$.

Local COPs are commonly pictured using *factor graphs* (see Fig. 1.4), *i.e.* bipartite graphs whose vertices are separated into *variable vertices*, which are associated to one of the variables x_i of the problem, and *factor vertices* which are associated to one of the constraints f_j. Each edge in the graph connects one factor vertex to one variable vertex, according to the rule that a factor $f_i(x_{j_1}, \ldots, x_{j_k})$ is connected to all and only the variables $x_{j_1}, \ldots, x_{j_k}$ on which it depends.

1.2.2 The Adiabatic Theorem

Adiabatic quantum computation is based on the adiabatic theorem, a result that intuitively states that the unitary dynamics generated by a time-dependent Hamiltonian $H(t)$ approximately preserve the istantaneous eigenspaces of the Hamiltonian at future times t, provided that the time derivative of the Hamiltonian is small enough. From this result follows in particular that a system starting in the ground state of a Hamiltonian $H(0)$ and evolving through a

slowly-changing Hamiltonian $H(t)$ stays in the (approximate) instantaneous ground state of $H(t)$ at every instant t of the evolution, provided this evolution is contained in some not-too-large time interval $[0, \tau]$.

Most of the literature points to [22] as the standard reference for the proof of the adiabatic theorem, even though other proofs are available [23] (the original proof is in [24]). We follow essentially the proof sketch given in [25]. Our starting point is the time-dependent Schrödinger Equation

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H(t)|\psi(t)\rangle$$

that we use to describe the evolution of a quantum system from time $t = 0$ to a final time $t = \tau$. By taking $\hbar = 1$ and changing the time t to a dimensionless time parameter $s = t/\tau$ we get

$$i\frac{\partial}{\partial s}|\psi(s)\rangle = \tau H(s)|\psi(s)\rangle \qquad \text{with } s \in [0, 1]. \tag{1.2}$$

Let $\{|n(s)\rangle\}_n$ be the (non-degenerate) eigenstates of the Hamiltonian $H(s)$, so that

$$H(s)|n(s)\rangle = E_n(s)|n(s)\rangle \qquad n = 0, 1, 2, \ldots \tag{1.3}$$

We formulate the following Ansatz for the solution of Eq. (1.2)

$$|\psi(s)\rangle = \sum_m \psi_m(s) e^{-i\tau\int_0^s E_m(s')\,\mathrm{d}s'}|m(s)\rangle,$$

where $\psi_n(s)$ are the variational parameters (actually, functions) we need to find. In foresight, we include the factors $\exp(-i\tau\int_0^s E_m(s')\,\mathrm{d}s')$ so as to cancel the dynamical phases of the propagation. The first step in proving the theorem is to derive the dynamical equations for $\psi_n(s)$. By multiplying both sides of Eq. (1.2) by $\langle n(s)|$ we obtain

$$i\langle n(s)|\frac{\partial}{\partial s}|\psi(s)\rangle = \tau\langle n(s)|H(s)|\psi(s)\rangle. \tag{1.4}$$

We compute both sides of this equation. If we take the ∂_s derivative of the Ansatz wavefunction we have

$$\begin{aligned}\frac{\partial}{\partial s}|\psi(s)\rangle &= \sum_m \frac{\partial\psi_m(s)}{\partial s} e^{-i\tau\int_0^s E_m(s')\,\mathrm{d}s'}|m(s)\rangle \\ &+ \sum_m \psi_m(s)(-i\tau E_m(s)) e^{-i\tau\int_0^s E_m(s')\,\mathrm{d}s'}|m(s)\rangle \\ &+ \sum_m \psi_m(s) e^{-i\tau\int_0^s E_m(s')\,\mathrm{d}s'}\frac{\partial}{\partial s}|m(s)\rangle,\end{aligned}$$

so that

$$\begin{aligned}i\langle n(s)|\frac{\partial}{\partial s}|\psi(s)\rangle &= i\frac{\partial\psi_n(s)}{\partial s} e^{-i\tau\int_0^s E_n(s')\,\mathrm{d}s'} \\ &+ \psi_n(s)\tau E_n(s) e^{-i\tau\int_0^s E_n(s')\,\mathrm{d}s'} \\ &+ i\sum_m \psi_m(s) e^{-i\tau\int_0^s E_m(s')\,\mathrm{d}s'}\langle n(s)|\frac{\partial}{\partial s}|m(s)\rangle.\end{aligned}$$

Moreover,

$$H(s)|\psi(s)\rangle = \sum_m \psi_m(s) e^{-i\tau \int_0^s E_m(s')\,\mathrm{d}s'} E_m(s)|m(s)\rangle$$

and therefore

$$\tau\langle n(s)|H(s)|\psi(s)\rangle = \psi_n(s) e^{-i\tau \int_0^s E_n(s')\,\mathrm{d}s'} \tau E_n(s).$$

Then from Eq. (1.4) we can write down the evolution equation for the coefficients $\psi_n(s)$ of the Ansatz

$$\begin{aligned}\frac{\partial \psi_n(s)}{\partial s} &= -\sum_m \psi_m(s) e^{-i\tau \int_0^s E_m(s) - E_n(s')\,\mathrm{d}s'} \langle n(s)|\frac{\partial}{\partial s}|m(s)\rangle \\ &= -\psi_n(s)\langle n(s)|\frac{\partial}{\partial s}|n(s)\rangle - \sum_{m\neq n} \psi_m(s) e^{-i\tau \int_0^s E_m(s) - E_n(s')\,\mathrm{d}s'} \langle n(s)|\frac{\partial}{\partial s}|m(s)\rangle\end{aligned}$$

For clarity we introduce the notation $\omega_{n,m}(s) = E_n(s) - E_m(s)$ and $F_{n,m}(s) = \langle n(s)|\partial H(s)/\partial s|m(s)\rangle$. Note that we always have the freedom to redefine the phase of the energy eigenstates $|n(s)\rangle \mapsto e^{i\theta_n(s)}|n(s)\rangle$. Where $\theta_n(s)$ is a continuous function of s. It is possible to choose $\theta_n(s)$ in such a way that by taking the derivative of the eigenvalue equation (1.3) one gets that

$$F_{n,m}(s) = \langle n(s)|\frac{\partial}{\partial s}|m(s)\rangle = \begin{cases} 0 & \text{if } n = m \\ \frac{\langle n(s)|\dot{H}(s)|m(s)\rangle}{E_n(s) - E_m(s)} & \text{otherwise} \end{cases},$$

where we abbreviate $\dot{H}(s) \equiv \partial H(s)/\partial s$. Therefore the s-derivative of the coefficients $\psi_n(s)$ simplifies to

$$\frac{\partial \psi_n(s)}{\partial s} = -\sum_{m\neq n} \psi_m(s) e^{-i\tau \int_0^s \omega_{m,n}(s')\,\mathrm{d}s'} F_{n,m}(s).$$

Integrating both sides from zero to s and imposing the initial condition $\psi_n(0) = \delta_{n,0}$, one gets

$$\psi_n(s) = \delta_{n,0} - \sum_{m\neq n} \int_0^s \psi_m(s) e^{-i\tau \int_0^{s'} \omega_{m,n}(s'')\,\mathrm{d}s''} F_{m,n}(s')\,\mathrm{d}s',$$

which gives the coefficient $\psi_n(s)$ as a sum of oscillating exponentials. Naively, one would expect that the largest contribution comes from $m = 0$ term in the sum, since the system starts its evolution in the ground state, which means that $\psi_0(s) = 1$ and $\psi_n(s) = 0$ for $n \neq 0$ at $s = 0$. We write

$$\psi_n(s) = -\int_0^s e^{-i\tau \int_0^{s'} \omega_{n,0}(s'')\,\mathrm{d}s''} F_{n,0}(s')\psi_0(s')\,\mathrm{d}s' + \text{other terms}.$$

Assuming the other terms can be neglected, we integrate by parts the first term

$$\psi_n(s) = -\frac{i}{\tau} e^{i\tau \int_0^{s'} \omega_{n,0}(s'')\,\mathrm{d}s''} \frac{F_{n,0}(s')}{\omega_{n,0}(s')}\psi_0(s')\Big|_0^s + \frac{i}{\tau}\int_0^s e^{i\tau \int_0^{s'} \omega_{n,0}(s'')} \,\mathrm{d}s' \frac{\partial}{\partial s}\Big(\frac{F_{n,0}(s')}{\omega_{n,0}(s')}\psi_0(s')\Big). \tag{1.5}$$

Now, in an ideal adiabatic evolution, the system remains in the ground state at all times s, so $\psi_n(s) = 0$ for all $n \neq 0$. It seems reasonable to ask that in an approximately adiabatic evolution all the components $\psi_n(s)$ with $n \neq 0$ should be small. The idea is then to assume that $|\psi_n(s)| \ll 1$ for all $n \neq 0$, and derive the conditions this assumption entails. We impose this smallness constraint on both terms in Eq. (1.5). Imposing the smallness of the first term one gets

$$1 \gg \frac{1}{\tau}\left|\frac{F_{n,0}(s)}{\omega_{n,0}(s)}\right|$$

Therefore, the time required to stay close to the adiabatic approximation, which we call the *adiabatic time* τ_a must satisfy

$$\tau_a \gg \frac{|\langle n(s)|\dot{H}(s)|0(s)\rangle|}{(E_n(s) - E_0(s))^2} \qquad \text{for all } 0 \leq s \leq 1, \tag{1.6}$$

which in turn implies (by taking $n = 1$) the more common *adiabatic condition*

$$\tau_a \gg \max_s \frac{|\langle 1(s)|\dot{H}(s)|0(s)\rangle|}{g(s)^2}, \tag{1.7}$$

where $g(s) \equiv \omega_{1,0}(s) = E_1(s) - E_0(s)$ is the gap separating the ground state and the first excited state of the Hamiltonian $H(s)$. This particular derivation of Eq. (1.7) presents it as a *necessary condition* for adiabaticity, *i.e.* for the system to stay in the ground state at all times of the evolution. Whether this is also a *sufficient condition* is a matter of some dispute. In this thesis however we will follow the common physicists' habit of considering Eq. (1.7) a *necessary and sufficient condition* for adiabaticity. We mention in passing that more stringent and more rigorous estimates of the error can be obtained (see *e.g.* [26, 27]), even though the formulas are more complicated.

1.2.3 Quantum Adiabatic Algorithm

The Adiabatic Quantum Algorithm's goal is to find the solution to a COP and its implementation proceeds in the following way. Given an instance of a COP described by a cost function f defined over N Boolean variables $X_1, \ldots, X_N$, one needs to define a Hamiltonian H_P in such a way that the ground state of H_P encodes the solution to the minimization problem described by f. A common choice is to take the a system of N quantum spins with Hilbert space $\mathcal{H} = \mathbb{C}^{2^N}$ and choose the N-fold σ^z-basis as the "computational basis". Then one labels each computational basis vector with a binary string through the definition $|0\rangle \equiv |\uparrow\rangle$ and $|1\rangle \equiv |\downarrow\rangle$. Then the encoding Hamiltonian is given by $H_P \equiv \sum_{x\in\{0,1\}^N} f(x)|x\rangle\langle x|$, which is clearly diagonal in the computational basis:

$$H_P = \begin{bmatrix} f(x_1) & 0 & & \ldots & 0 \\ 0 & f(x_2) & 0 & \ldots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & & \ldots & 0 & f(x_{2^N}) \end{bmatrix}.$$

One then defines an *initial Hamiltonian* H_0 (defined in the same Hilbert space $\mathcal{H}$ over which H_P acts), a total annealing time τ, and a regular (at least C^1) parametrized curve $H : [0, \tau] \to \mathrm{Herm}(\mathcal{H})$ in the space of self-adjoint operators over $\mathcal{H}$, with the following properties

(i) $H(0) = H_0$ and $H(\tau) = H_P$, that is the curve connects the Hamiltonian H_0 to the Hamiltonian H_P

(ii) a ground state of H_0 is "easy to prepare". This is a technical requirement but we will take it to mean that H_0 must be a non-interacting Hamiltonian. In an experimental implementation this ground state is usually generated by cooling the system down to a very small temperature (ideally $T = 0$).

The adiabatic algorithm then proceeds as follows.

1. Prepare the system in the ground state $|\psi_0\rangle$ of the initial Hamiltonian H_0.
2. Let the system evolve for a time τ under the dynamics defined by the time-dependent Hamiltonian $H(t)$.
3. Measure the system in the computational basis.

The adiabatic theorem then guarantees that if $\tau \geq \tau_a$ (the adiabatic time defined in the previous section), then the state of the system at time $t = \tau$ will have a constant overlap with the ground state of the problem Hamiltonian H_P, so the measurement will return the optimal solution to the CPS with high probability.

Let us restrict immediately to a concrete example, that is going to apply for the rest of this thesis. For a given problem Hamiltonian $H_P \in \mathrm{Herm}(\mathcal{H})$ we choose the initial Hamiltonian to be the transverse-field term

$$H_0 = -\Gamma \sum_i \sigma_i^x, \tag{1.8}$$

where $\Gamma > 0$ is a parameter that tunes its intensity. The path connecting the two Hamiltonians is just the straight line segment

$$\gamma = \{\alpha H_0 + (1-\alpha) H_P \mid 0 \leq \alpha \leq 1, \}$$

which we follow at constant speed $v \equiv 1/\tau$:

$$H(t) = \frac{\tau - t}{\tau} H_0 + \frac{t}{\tau} H_P.$$

Now let us see what this entails for the adiabatic condition (1.7). Note that with this choice we have $\dot{H}(s) = H_P - H_0$ and then

$$\begin{aligned} \left|\langle 1(s)|\dot{H}(s)|0(s)\rangle\right| &\leq \left|\langle 1(s)|H_P|0(s)\rangle\right| + \left|\langle 1(s)|H_0|0(s)\rangle\right| && (1.9)\\ &\leq \|H_P\|^{1/2} + \|H_0\|^{1/2} && (1.10)\\ &= |\lambda_{\max}^{(P)}|^{1/4} + |\lambda_{\max}^{(0)}|^{1/4}. && (1.11) \end{aligned}$$

Where $\lambda_{\max}^{(P)}$ and $\lambda_{\max}^{(0)}$ are the largest eigenvalues of H_P and H_0, respectively. In the derivation we have used the fact that if $|\varphi\rangle, |\psi\rangle$ are normalized vectors, then $|\langle\varphi|A|\psi\rangle|^2 \leq \|A|\psi\rangle\|^2 \leq \|A\|^2$ and that the operator norm $\|A\|$ induced by an ℓ_2-norm is given by the maximal singular value of A.

If H_P encodes a COP and H_0 is a transverse field in Eq. (1.8) then we see that $\lambda_{\max}^{(0)} = \Gamma N$ and $\lambda_{\max}^{(0)}$ is of polynomial order in the size of the system N, because H_P has a discrete spectrum given by sum of the costs that the constraints associate to a binary assigment. Since by definition a COP has

a number of constraints that is polynomial in the number of variables, and we assumed that the costs are $O(1)$ in N, then $\lambda_{\max}^{(P)} = O(\mathsf{poly}(N))$ and the numerator in the adiabatic condition (1.7) is of polynomial order in N, and independent of the dimensionless time parameter s:

$$\Big|\langle 1(s)|\dot{H}(s)|0(s)\rangle\Big| = O(\mathsf{poly}(N)).$$

The adiabatic time τ is then a function of the square of the inverse gap $g(s)^2$, as one has to maximize $1/g(s)^2$ for $0 \leq s \leq 1$. The solution is fixed by the minimal gap $g_{\min}$ of $H(s)$ in the adiabatic path.

The asymptotic runtime of the adiabatic algorithm is determined by how $g_{\min}$ varies as a function of N. In particular, if $g_{\min} \sim 1/\mathsf{poly}(N)$ in the limit of $N \to \infty$, then $\tau(N) = O(\mathsf{poly}(N))$ and the adiabatic algorithm is *efficient*. If, on the other hand, $g_{\min} \sim 1/\exp(N)$ then $\tau(N) = \Omega(\exp(N))$ and the algorithm is not efficient.

1.3 Physical Interpretations

The adiabatic formula Eq. (1.7) gives us the conditions under which an annealing schedule may be considered adiabatic. These conditions however involve spectral gaps and matrix elements that are usually hard to compute and perhaps a bit opaque in their physical interpretation. It seems therefore useful to develop some different physical intuition in order to better understand how the quantum adiabatic algorithm works. Indeed, there have been a few proposals and we are going to consider them in this section.

1.3.1 Tunnelling through energy barriers

In the seminal paper where quantum annealing was introduced, the authors considered an adiabatic path of the form

$$H(t) = H_P - \Gamma(t) \sum_i \sigma_i^x,$$

where $\Gamma(t)$ is decreased to zero from a inital value $\Gamma(0) > 0$ and H_P is the problem Hamiltonian which is taken to be diagonal in the computational σ^z-basis. One can see that in this case the transverse field term induces transitions between classical energy configurations (*i.e.* computational basis states): for $t \ll 1$ and to leading order in Γ:

$$\begin{aligned} |\langle y|U(t)|y'\rangle| &\approx |\langle y|e^{it\Gamma \sum_j \sigma_j^x} e^{-itH_P}|y'\rangle| = |e^{-itE(y')}||\langle y|e^{it\Gamma \sum_j \sigma_j^x}|y'\rangle| \\ &= |\langle y|e^{it\Gamma \sum_j \sigma_j^x}|y'\rangle| = O\Big(\frac{t^k \Gamma^k}{k!}\Big), \end{aligned}$$

where y, y' are labels for states in the computational (*i.e.* σ^z diagonal) basis and $k \equiv \mathrm{dist}(y, y')$ is the Hamming distance[3] between y and y'. The idea is then to

[3] given two strings $x_1, \ldots, x_N$ and $y_1, \ldots, y_N$, the Hamming distance between them is the number of positions i where $x_i \neq y_i$. By extension, if we have two quantum product states $\bigotimes_i |x_i\rangle$ and $\bigotimes_i |y_i\rangle$ then we can define a Hamming distance between them by the number of sites i where $|x_i\rangle \neq |y_i\rangle$.

start in a state that is a uniform superposition of all classical energy eigenstates (*i.e.* the ground state of the transverse field term, a state with large quantum fluctuations) and successively reduce the strength of the transverse field. Then the amplitudes of the (classical) excited states will escape their "metastable state" through quantum tunnelling and will concentrate in the ground state of the classical problem. If the decrease of Γ is slow enough, one can prove [28] that in the infinite-time limit the system will reach the ground state with probability one.

Accepting the "tunnelling between classical states" pricture, one would like to derive some ballpark estimate of the efficiency of such process. The usual benchmark against which this process is tested is simulated annealing (see Appendix A.2 for the details), on the argument that both algorithms are non-deterministic algorithms that use physical processes to explore an exponentially-large configuration space. One way to do that is to use the Wentzen-Kramers-Brillouin approximation to study the expected tunnelling time of a particle with mass m through an energy barrier, which is the supposed mechanism used by the quantum annealing procedure to overcome the energy barriers in the potential according to this interpretation. In the case of one-dimensional barrier of height ΔE and width w, one gets

$$\tau = e^{w\sqrt{\frac{1}{2}m\Delta E}}.$$

By comparison, the Kramers escape problem describes the expected escape time from a metastable state with a barrier of height ΔE through thermal excitations. This is used to model the dynamics by which thermal annealing (and its simulated version) can find its way through the energy landscape and reach the ground state. It solution gives an Arrhenius-like time of

$$\tau = e^{\Delta E/T}.$$

Then the argument is that quantum annealing should outperform classical thermal annealing in problems where the metastable states in the energy landscape are separated by tall but thin barriers (of width $w = O(1)$), so that the thermal escape goes like $\exp(\Delta E)$ while the tunnelling time goes like $\exp(\sqrt{\Delta E})$. Note that this (admittedly very crude) estimate entails a quadratic speedup in an algorithm that runs in exponential time, *i.e.* a Grover-like speedup [29]. Unfortunately, at the present time no realistic combinatorial problem is known to have this kind of energy landscape and in fact, these properties of energy landscapes are very difficult to study.

However appealing on an intuitive level, this picture suffers from a few problems. An obvious one is that escape times are calculated for one-dimensional potentials while in real combinatorial problems one has highly multidimensional Hilbert spaces, so the comparison presented above is at best a cartoon of the true setup. More importantly, it seems hard to completely trust this particular tunnelling picture or at least to use it for quantitative analysis. This is because in order to be able to speak sensibly about tunnelling one requires a semiclassical potential to define classically allowed and forbidden regions. The recipe of using the classical energy landscape cannot be correct in the large-Γ limit, where the energy of the system is dominated by the transverse-field term. Here we have a single ground state separated by a gap $g \approx 2\Gamma$ from the rest of the spectrum. One expects to start finding a corrugated energy landscape with a

complicated set of metastable states when the system enters the glassy phase. Nevertheless, quantum chaos in the Γ parameter will likely shift metastable states up and down as Γ is decreased so that the energy landscape inside of the glassy phase will look nothing like the classical one until the very end of the annealing process. At this point the spread of the wavefuction over the classical states will strongly depend on the previous part of the annealing protocol, which this picture does not capture.

1.3.2 Landau-Zener processes

The two seminal papers [30, 31] by Lev Landau and Clarenece Zener describe a quasi-adiabatic process where the time-dependent Hamiltonian $H(t)$ of a two-level system has two energy levels that start infinitely far apart at $t = -\infty$ are brought close together at $t = 0$, where the gap between them is $2\mu > 0$, and are then separated again in the limit of $t = \infty$ (see Fig. 1.5). Assuming the system starts in the ground state at time $t = -\infty$, how is the amplitude of the wavefunction spread between the two levels at time $t = \infty$?

The system's Hamiltonian can be written using Pauli matrices: $H(t) = \mu\sigma^x + \lambda(t)\sigma^z$. In the diabatic ($\sigma^z$-diagonal) basis $\{|0\rangle, |1\rangle\}$ this is the 2×2 matrix

$$H(t) = \begin{bmatrix} \lambda(t) & \mu \\ \mu & -\lambda(t) \end{bmatrix}, \qquad \text{where } \lambda(t) = vt \text{ for } t \in \mathbb{R}.$$

The parameter $v \equiv v_{LZ} > 0$ is the Landau-Zener velocity and describes how quickly the Hamiltonian is changing with time. The instantaneous energy eigenstates of the system are

$$\begin{aligned} |E_0(t)\rangle &= \sin(\theta/2)|0\rangle - \cos(\theta/2)|1\rangle \\ |E_1(t)\rangle &= \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle, \end{aligned}$$

where $\theta = \arctan(\mu/\lambda(t))$. Note that in the limit $t \to \pm\infty$ these converge to the eigenvectors $|0\rangle, |1\rangle$ of σ^z, but in different order

$$\begin{aligned} |0\rangle &= |E_0(-\infty)\rangle = |E_1(\infty)\rangle \\ |1\rangle &= |E_1(-\infty)\rangle = |E_0(\infty)\rangle. \end{aligned}$$

These energy eigenstates $|E_0(t)\rangle, |E_1(t)\rangle$ are associated to the energy levels

$$\begin{aligned} E_0(t) &= -\sqrt{\mu^2 + \lambda(t)^2} \\ E_1(t) &= \sqrt{\mu^2 + \lambda(t)^2}, \end{aligned}$$

so the gap between the two levels is given by $g(t) = 2\sqrt{\mu^2 + \lambda(t)^2}$ and the minimal gap is $g_{\min} = 2\mu$. If the system starts in the ground state at time $t = -\infty$, then the probability of being in the excited state $|E_1(t)\rangle$ at time $t = \infty$ is given by

$$\Big|\langle E_1(\infty)|U(\infty, -\infty)|E_0(-\infty)\rangle\Big|^2 = \exp\Big(-\frac{\pi\mu}{v}\Big). \tag{1.12}$$

One can see that the nature of the ground state wavefunction changes abruptly from $|0\rangle$ to $|1\rangle$ around the avoided crossing, so in order to remain

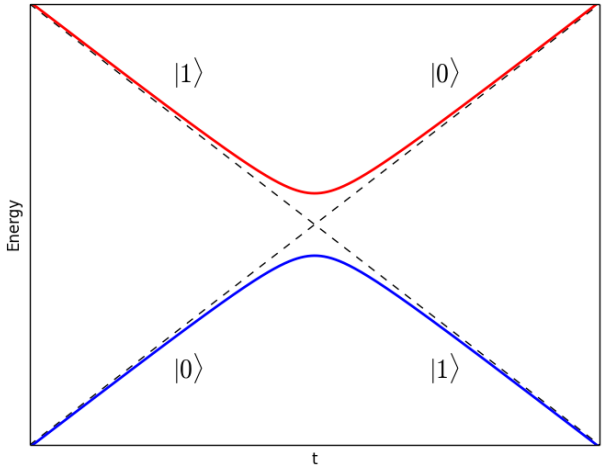


Figure 1.5: Pictorial representation of a Landau-Zener transition. The blue line is the energy of the ground state $|E_0(t)\rangle$ while the red one is the energy of the first excited state $|E_1(t)\rangle$. If $\mu > 0$ then the two levels never cross but have a minimal gap of 2μ at time $t = 0$.

in the ground state, the system has to transition quickly from $|0\rangle$ to $|1\rangle$. From Eq. (1.12) one can see that this transition is hampered by small minimal gap $g_{\min} = 2\mu$ and is eased by a small Landau-Zener velocity v.

Note the similarity between the adiabatic condition (1.7) and the Landau-Zener excitation probability (1.12). The excitation probability decreases exponentially as a function of the minimal gap $g_{\min}$, so that if we consider a family of Landau-Zener Hamiltonians $\{H_n(g_{\min}^{(n)})\}_n$ where the minimal gap is closing in the limit $g_{\min}^{(n)} \xrightarrow{n} 0$, then the probability of staying in the ground state vanishes as n approaches infinity (*i.e.* the transition becomes less and less adiabatic). This effect needs to be countered by decreasing the Landau-Zener velocity v, which can be taken to be the analogue of having an increasingly large annealing time τ in the adiabatic algorithm setup. This has led people into modelling the physical process of the adiabatic quantum algorithm as a cascade of Landau-Zener anticrossings happening at different times and in different regions of the spectrum. Of course care must be taken when doing this, as standard Landau-Zener process is a two-level process while in a run of the adiabatic algorithm we need to consider the full spectrum of the Hamiltonian, so everthing must be taken as an approximation. In particular, this approximation seems most reasonable when the affected energy levels are well-separated from the rest of the spectrum by a substantial energy gap. It is most definitely *not* a good approximation when the system undergoes a second-order phase transition and the minimal gap closes with a continuous spectrum on top of the ground state.

1.3.3 Path-integral Representation

Path integrals are ubiquitous in quantum mechanics. In this section we write the real-time propagator of the time-dependent Hamiltonian $H(t)$ in terms of path integrals:

$$U(1,0) = \sum_{n,m} \Big(\int \mathcal{D}[p]\, W(p) \Big) |n(1)\rangle\langle m(0)|, \tag{1.13}$$

where $\mathcal{D}[p]$ is the measure over the set of paths $p : [0,1] \to \mathrm{Sp}(H(s))$ through the istantaneous energy eigenstates, $W(p)$ is the (generically complex) weight associated to the path p and $|n(s)\rangle$ are the eigenstates of the Hamiltonian $H(s)$. We believe the path-integral representation has the advantage of making plain the fact that the adiabatic theorem is at its core an *interference phenomenon* where with decreasing adiabatic velocity different paths that would lead the system to an excited state interfere destructively.

We follow the presentation in [32]. We start from the time-dependent Scrödinger Equation in dimensionless time $s = t/\tau \in [0,1]$:

$$i\frac{\partial}{\partial s}|\psi(s)\rangle = \tau H(s)|\psi(s)\rangle.$$

The unitary propagator $U(s,0)$ generated by this equation is given by the time-ordered exponential

$$U(s,0) = \mathbb{T} \exp\Big(-i\tau \int_0^s H(s')\, \mathrm{d}s' \Big),$$

where $\mathbb{T}$ is the time-ordering operator.

We can use the semigroup property of the propagator $U(s'',s) = U(s'',s')U(s',s)$ to split the evolution interval $[0,1]$ into N equal subintervals, $\Delta s \equiv 1/N$

$$0 \equiv s_0 < s_1 < \ldots < s_N \equiv 1,$$

where $s_k \equiv k\Delta s$, so that

$$U(s) = U_{N-1} \cdots U_1 U_0, \tag{1.14}$$

where $U_k \equiv U(s_{k+1}, s_k)$ and we take the approximation $U(s_{k+1}, s_k) \approx e^{-i\tau H(s_k)\Delta s}$. This approximation is exact in the limit $N \to \infty$.

The usual way to define a path integral is to insert resolutions of the identity between successive terms $U_{k+1}U_k$, so that the matrix element of the propagator $\langle y|U(s)|x\rangle$ can be written as a sum of products of the matrix elements of the short-time propagators U_k

$$\langle y|U(s)|x\rangle = \sum_{x^0 \cdots x^{(N-2)}} \langle y|U_{N-1}|x^{(N-2)}\rangle \cdots \langle x^{(1)}|U_1|x^{(0)}\rangle\langle x^{(0)}|U_0|x\rangle$$

and then take the limit $\Delta s \to 0$. The first thing we need to do is to compute the matrix element of a propagator for the $\langle y|U_0|x\rangle$. We can expand $U(s_{k+1}, s_k)$ as

$$U_k(s_{k+1}, s_k) = \mathbb{I} - i\tau\Delta s H(s_k) + O\Big(\Delta s^2\Big).$$

Notice that is we bracket this propagator with an energy eigenvector $|m(s_{k+1})\rangle$ of $H(s_{k+1})$ on the left, and an energy eigenvector $|n(s_k)\rangle$ of $H(s_k)$ on the right, we obtain

$$\langle m(s_{k+1})|U_k(s_{k+1},s_k)|n(s_k)\rangle = \langle m(s_{k+1})|n(s_k)\rangle - i\tau E_n(s_k)\Delta s\langle m(s_{k+1})|n(s_k)\rangle + O\Big(\Delta s^2\Big).$$

By expanding $\langle m(s)|$ around $s = s_k$ we get

$$\langle m(s_{k+1})| = \langle m(s_k)| + \Delta s\langle \dot{m}(s_k)| + O\Big(\Delta s^2\Big)$$

and by plugging this in the previous equation we obtain

$$\begin{aligned}\langle m(s_{k+1})|U_k(s_{k+1},s_k)|n(s_k)\rangle &= \delta_{mn}\Big(\mathbb{I} - i\tau E_n(s_k)\Delta s\Big) + \Delta s\langle \dot{m}(s_k)|n(s_k)\rangle + O\Big(\Delta s^2\Big)\\ &= \delta_{mn}e^{-i\tau E_n(s_k)\Delta s} + \Delta s\langle \dot{m}(s_k)|n(s_k)\rangle + O\Big(\Delta s^2\Big),\end{aligned}$$

since $\mathbb{I} + \alpha A = \exp(\alpha A) + O(\alpha^2)$. Now we insert resolutions of the identity $\mathbb{I} = \sum_n |n(s)\rangle\langle n(s)|$, where $|n(s)\rangle$ is the eigenbasis of $H(s)$, between the short-time propagators U_k in Eq. (1.14). We get

$$\begin{aligned}\langle m(s)|U(s)|m_0(0)\rangle &= \sum_{m_{N-1}}\cdots\sum_{m_1}\langle m(s)|U_{N-1}|m_{N-1}(s_{N-1})\rangle\cdots\langle m_1(s_1)|U_0|m_0(0)\rangle\\ &= \sum_{m_{N-1}}\cdots\sum_{m_1}\prod \delta_{mn}e^{-i\tau E_n(s_k)\Delta s} + \Delta s\langle \dot{m}(s_k)|n(s_k)\rangle + O\Big(\Delta s^2\Big).\end{aligned}$$

This is essentially the path-integral formula of Eq. (1.13), one only needs to rearrange the terms to make its form explicit. We organize the terms of this sum by the number $N_{\Delta s}$ of Δs factors it contains. For $N_{\Delta s} = 0$ the only non-zero terms are of the form

$$e^{-i\tau\sum_k E_n(s_k)\Delta s} \qquad \text{for fixed } n \in \mathbb{N},$$

that in the limit $\Delta s \to 0$ become $e^{-i\tau\int_0^1 E_n(s')\,\mathrm{d}s'}$. This can be seen pictorially as a path starting in the state $|n(0)\rangle$ and evolving to the state $|n(s)\rangle$ for all times $0 \le s \le 1$. During this evolution, the state of the system accumulates only a dynamical phase. For $N_{\Delta s} = 1$ we have a sum of terms of the form

$$\sum_K e^{-i\tau\sum_{k=0}^K E_n(s_k)\Delta s}\Delta s\langle \dot{m}(s_k)|n(s_k)\rangle e^{-i\tau\sum_{k=K}^M E_m(s_k)\Delta s}$$

for fixed $n, m \in \mathbb{N}$. In the limit $\Delta s \to 0$ these become

$$\int_0^1 e^{-i\tau\int_0^{s_0} E_n(s')\,\mathrm{d}s'}\langle \dot{m}(s_0)|n(s_0)\rangle e^{-i\tau\int_{s_0}^1 E_m(s')\,\mathrm{d}s'}\,\mathrm{d}s_0.$$

Notice that the integrand (for a fixed value of s_0) is the contribution of a path that starts in the state $|n(0)\rangle$, evolves in the state $|0(s)\rangle$ for all $s < s_0$, then jumps to $|m(s_0)\rangle$ at time $s = s_0$, and then evolves to $|m(s)\rangle$ for all $s > s_0$. This jump contributes with a factor $\langle \dot{m}(s_0)|n(s_0)\rangle$ to the weight of the path.

More generally, each of the non-zero terms (in the limit $\Delta s \to 0$) in the sum can be associated to a path, *i.e.* a specific realization of a continuous-time jump

process that moves between the istantaneous eigenstates of the system. A path is a function $p : [0,1] \to \mathbb{N}$ with the intended meaning that the index $p(s) \in \mathbb{N}$ labels an eigenstate of the Hamiltonian $H(s)$, *e.g.* $p(s) = k$ means that the system at time s is in the eigenstate $|k(s)\rangle$ of $H(s)$. The weight $W(p)$ of the path p is given by

$$W(p) = e^{-i\tau \int_0^{s_1} E_{n_1}(s)\,\mathrm{d}s + \cdots + \int_{s_k}^1 E_{n_k}(s)\,\mathrm{d}s} \langle \dot{n}_1(s_1)|n_2(s_1)\rangle \cdots \langle \dot{n}_k(s_k)|n_{k+1}(s_k)\rangle,$$

where one can make a choice of the phases $e^{i\phi_n(s)}$ for the eigenstates of $H(s)$ so that

$$\langle \dot{n}(s)|m(s)\rangle = \frac{\langle n(s)|\dot{H}(s)|m(s)\rangle}{E_n(s) - E_m(s)} = F_{n,m}(s).$$

Note that the weights $W(p)$ are not necessarily phases: in addition to the dynamical phases we have the factors $\langle \dot{n}(s)|m(s)\rangle$ associated to the jumps. These can have arbitrary magnitude. This means, for example, that if we assume the existence of a typical value for the matrix elements $\langle n(s)|\dot{H}(s)|m(s)\rangle \approx V_{\text{typical}}$, we see that paths that have either (i) a large number of jumps, or (ii) jumps between states with large differences in energy, will give a small contribution to the integral (1.13).

If $U(1,0)$ is the propagator of a run of the quantum adiabatic algorithm, then the final state is $U(1,0)|0\rangle$. The contribution of all the one-jump paths that start in the ground state and jump to an excited $|m(s)\rangle$ state at some time s can be written as

$$e^{-i\tau \int_0^1 E_m(s')\,\mathrm{d}s'} \int_0^1 e^{-i\tau \int_0^s \omega_{0,n}(s')\,\mathrm{d}s'} \langle \dot{m}(s)|0(s)\rangle\,\mathrm{d}s$$

where $\omega_{0,n}(s') = E_0(s') - E_m(s')$ as in Section 1.2.2. If we only consider the contributions given by the paths with exactly one jump, then on gets the error term is approximated by [27]

$$\Big\| \Big(\mathbb{I} - |0(1)\rangle\langle 0(1)| \Big) U(1,0)|0\rangle \Big\|^2 \approx \sum_{m \neq 0} \Big| \int_0^1 F_{m,0}(s) e^{-i\tau \int_0^s \omega_{0,n}(s')\,\mathrm{d}s'}\,\mathrm{d}s \Big|^2.$$

and integrating by parts the intergral in the l.h.s term, we get

$$\frac{F_{m,0}(s)}{-i\tau\omega_{0,n}(s)} e^{-i\tau \int_0^s \omega_{0,n}(s')\,\mathrm{d}s'} \Big|_0^1 - \int_0^1 \Big(\frac{\partial}{\partial s} \Big[\frac{F_{m,0}(s)}{-i\tau\omega_{0,n}(s)} \Big] e^{-i\tau \int_0^s \omega_{0,n}(s')\,\mathrm{d}s'} \Big).$$

One can prove that the norm of this term can be upper bounded so as to get

$$\Big\| \sum_{m \neq 0} |m(1)\rangle \int_0^1 F_{m,0}(s) e^{-i\tau \int_0^s \omega_{0,n}(s')\,\mathrm{d}s'}\,\mathrm{d}s \Big\| \leq 2 \frac{\max_s \|\dot{H}(s)\|}{\min_s g(s)^2 \tau} + O(\tau^{-2}),$$

which is essentially Eq. (1.7). One then sees that the usual adiabatic condition is equivalent to neglecting the contributions given by the paths having more than one jump.

1.4 Entanglement and Adiabatic Computation

Entanglement is one of the principal features that are specifically peculiar to composite quantum mechanical systems, as was discovered early on in the history of quantum mechanics. Its presence signifies the existence of non-local correlations between (possibly remote, spatially separated) degrees of freedom, to the effect that the state of the composite system cannot be reconstructed from the local states of its components. In mathematical terms, this means that the pure state $|\psi\rangle$ of a many-body system with Hilbert space $\mathcal{H} = \bigotimes_i \mathcal{H}_i$ does not factorize over the tensor product structure of $\mathcal{H}$

$$|\psi\rangle \neq |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_N\rangle.$$

While the factorizable/non factorizable question describes a binary property, a great effort went into trying to find quantitative measures for the different ways in which a state can be entangled. This gave rise to a vast number of entanglement measures, each of them designed to capture a specific consequence for a state to be non-factorizable.

In this thesis we will be mainly concerned with three entanglement measures for pure states: the Von Neumann entanglement entropy, the Rényi entanglement entropies and the quantum Fisher information. The first two quantify *bipartite* entanglement, meaning that they depend on a specific factorization of the many-body Hilbert space in two factors $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ and measure only the quantum correlations *across* this partition. Given this choice and a pure state $|\psi\rangle$ one computes the reduced density matrix ρ_A of system A by tracing away from the $\rho = |\psi\rangle\langle\psi|$ the degrees of freedom belonging the system B

$$\rho_A \equiv \mathrm{Tr}_B(\rho) = \mathrm{Tr}_B(|\psi\rangle\langle\psi|),$$

and then computes a function $f(\rho_A)$ of the reduced density matrix that vanishes on matrices of rank one (that is, on pure states). The Von Neumann entanglement entropy $S(\rho_A)$ is defined by taking this function to be $f(X) = -\mathrm{Tr}(X \log X)$:

$$S(\rho_A) \equiv -\mathrm{Tr}\left(\rho_a \log(\rho_A)\right).$$

Equivalently, this means computing the Shannon entropy of the (discrete, since we are always finite-dimensional) spectrum of ρ_A. If $\mathrm{Sp}(\rho_A) = \{\lambda_1, \ldots, \lambda_N\}$ then

$$S(\rho_A) = -\sum_{i=1}^{N} \lambda_i \log(\lambda_i).$$

Second among the entanglement measures we will consider, the entanglement Rényi entropies $S^{(\alpha)}$ (α-Rényi entropy for short) are a family of entropy measures indicized by the real parameter $\alpha > 0$ and defined by the functions $f_\alpha(X) = \frac{1}{1-\alpha} \log \mathrm{Tr}(X^\alpha)$:

$$S^{(\alpha)}(\rho_A) = \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho_A^\alpha)$$

Equivalently, one needs to compute the (classical) Rényi entropy of the spectrum of ρ_A:

$$S^{(\alpha)}(\rho_A) = \frac{1}{1-\alpha} \log \sum_{i=1}^{N} \lambda_i^\alpha.$$

One can show that the α-Rényi entropy converges to the Von Neumann entropy in the limit $\alpha \to 1$:

$$\lim_{\alpha \to 1} S^{(\alpha)}(X) = S(X)$$

Bipartite entanglement has a role to play in adiabatic computing as recent numerical evidence [33, 34] suggests that bipartite entanglement positively correlates with better performances of a quantum annealer (*i.e.* better approximations to the ground state energy). This seems to indicate that at least during some parts of the adiabatic run, the system must inhabit (relatively) highly entangled states, and that by forcibly limiting the amount of entanglement that is available to the system, as was done in the cited works, one is effectively throwing the system off-equilibrium and into excited states (which are associated to less optimal solutions to the combinatorial problem).

The third entanglement measure we are going to consider is the quantum Fisher information F_Q. The quantum Fisher information was first used in quantum metrology to measure the efficacy of a phase-estimation protocol [35, 36, 37], when an initial state ρ of a many-body system is evolved as

$$\rho \mapsto e^{-i\theta H} \rho e^{i\theta H} \equiv \rho(\theta) \tag{1.15}$$

where H is a local Hamiltonian $H = \sum_i h_i$, with h_i acting nontrivially only on the i-th site, and $\theta \in \mathbb{R}$ is a fixed phase. We assume for simplicity that $\|H\| = 1$ but note that is not a crucial assumption for the results we are going to describe. One then would like to learn the value of θ by performing some measurement on (independent copies of) $\rho(\theta)$. This is accomplished first by measuring $\rho(\theta)$ with respect to a positive operator-valued measure (POVM) $\{E_\mu\}$, that gives the result μ, belonging to a set $\mathcal{M}$ of possible outcomes, with probability

$$p(\mu|\theta) \equiv \mathrm{Tr}(E_\mu \rho(\theta)).$$

After m repetitions of the measurement on independent copies of $\rho(\theta)$, one applies an *estimator* $\hat{\theta} : \mathcal{M}^m \to \mathbb{R}$ to the results $\mu_1, \ldots, \mu_m$ thus obtained, so that $\hat{\theta}(\mu_1, \ldots, \mu_m)$ is meant to represent the "guess" for the value of θ. Note that since the outcomes μ_i of the measurements are random, then $\hat{\theta} = \hat{\theta}(\mu_1, \ldots, \mu_m)$ is a random variable. Then the Fisher information F

$$F[\rho(\theta), \{E_\mu\}, H] = \sum_\mu \frac{1}{p(\mu|\theta)} \Big(\frac{\partial}{\partial \theta} p(\mu|\theta) \Big)^2$$

is a quantity (dependent on the Hamiltonian H, the initial state ρ, the phase θ and the choice of POVM $\{E_\mu\}$) that limits the accuracy of an unbiased[4] estimator $\hat{\theta}$ through the Cramer-Rao bound on its standard deviation

$$\sigma(\hat{\theta}) \geq \frac{1}{\sqrt{mF}}. \tag{1.16}$$

It turns out that in the case described by Eq. (1.15), the Fisher information does not depend on the value θ, so we can write it as $F[\rho, \{E_\mu\}, H]$. The optimization

[4] meaning that the expectation value of $\hat{\theta}$ gives the correct value θ for the parameter, $\big\langle \hat{\theta} \big\rangle = \theta$.

of this protocol with respect to the choice of POVM defines the *quantum Fisher information* F_Q (with respect to the generator H)

$$F_Q[\rho, H] \equiv \sup_{\{E_\mu\}} F[\rho, \{E_\mu\}, H], \tag{1.17}$$

which gives rise to the generalized version of Eq. (1.16)

$$\sigma(\hat{\theta}) \geq \frac{1}{\sqrt{mF}} \geq \frac{1}{\sqrt{mF_Q}}. \tag{1.18}$$

One can prove that the inequalities in the equation can be saturated by an optimal POVM $\{E_\mu\}$, so that the least upper bound in Eq. (1.17) can be taken to be a maximum. A consequence of this fact is that a phase-encoding procedure (1.15) with a large value of F_Q allows for better phase estimation.

For a pure state $\rho = |\psi\rangle\langle\psi|$, the quantum Fisher information takes a simple form

$$F_Q[\psi, H] = 4\,\mathrm{Var(H)} = 4\Big(\langle\psi|H^2|\psi\rangle - \langle\psi|H|\psi\rangle^2\Big). \tag{1.19}$$

If the state $|\psi\rangle$ is entangled, then the quantum Fisher information density f_Q satisfies the inequality

$$f_Q[\psi, H] \equiv \frac{F_Q[\psi, H]}{N} \leq k \tag{1.20}$$

where k is the size of the largest "entangled block" in the N-spin state $|\psi\rangle$, *i.e.* if the state $|\psi\rangle$ is written as the tensor product of states $|\psi_i\rangle$

$$|\psi\rangle = |\psi_1\rangle \otimes \cdots \otimes |\psi_m\rangle$$

where for every i the ket $|\psi\rangle_i$ is a state of n_i spins forming a (not necessarily contiguous) "block" of spins, then $k \equiv \max_i\{n_i\}$. One can then use the quantum Fisher information as an entanglement witness: if $|\psi\rangle$ contains an entangled block of size greater than k, then the inequality (1.20) is violated for some choice of phase shift generator H. By maximizing $F[\psi, H]$ over all possible non-interacting Hamiltonians $H = \sum_i h_i$ one can learn the size of the largest entangled block in $|\psi\rangle$. States whose maximal entangled block is of size k or less are called *k-producible* or *k-blocked* in the literature.

The kind of multipartite entanglement detected by the quantum Fisher information is also known to play a role in the theory of quantum computational complexity: volume-law entanglement scaling was observed to be associated with the *exponential* speedup that quantum computers are expected to have over classical computers at performing certain specific computational tasks [38]. On the contrary, it is known that efficient quantum algorithms that generate an entanglement that grows slowly with the system size (as measured *e.g.* by the Schmidt rank [39] or the size of the maximal entangled block [40]) can achieve at most a polynomial speedup over their classical counterparts.

1.4.1 Replica Method for the Rényi Entropy

We have seen that in order to compute the entanglement entropies of a generic spin system one needs either the reduced density matrix ρ_A of the state, or its

spectrum. This is in general a very resource-intensive task due to the exponential growth of the dimension of the Hilbert space as a function of the number of spins in the system. If ρ is the thermal state of a system with Hamiltonian H, *i.e.* $\rho \propto e^{-\beta H}$ and one is interested in computing only the α-Rényi entropies with integer order α, then there is an easier approach that uses the replica trick. For better clarity we describe the method in the case $\alpha = 2$ but this is easily generalized for any integer greater than one. First one writes ρ as a rank four tensor in a product basis $\{|a\rangle|b\rangle\}$ of $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$:

$$\rho_{b,b'}^{a,a'} = \langle a|\langle b|\rho|a'\rangle|b'\rangle = \frac{1}{Z}\langle a|\langle b|e^{-\beta H}|a'\rangle|b'\rangle.$$

Now it is easy to see that the matrix elements of $\rho_A = \mathrm{Tr}_B(\rho)$ can be written as

$$\langle a|\rho_A|a'\rangle = \sum_b \rho_{b,b}^{a,a'} = \frac{1}{Z}\sum_b \langle a|\langle b|e^{-\beta H}|a'\rangle|b\rangle$$

and $\mathrm{Tr}(\rho_A^2) = \mathrm{Tr}(\rho_A\rho_A)$ is given by

$$\mathrm{Tr}(\rho_A^2) = \sum_{a,a'}\langle a|\rho_A|a'\rangle\langle a'|\rho_A|a\rangle = \frac{1}{Z^2}\sum_{a,a',b,b'}\langle a|\langle b|e^{-\beta H}|a'\rangle|b\rangle\langle a'|\langle b'|e^{-\beta H}|a\rangle|b'\rangle. \tag{1.21}$$

Note that the sum in the equation above can be interpreted as the partition function $Z_A^{(2)}$ of a quantum mechanical system made of two replicas of the original system, where the region A is propagated for an imaginary time of 2β *across* the replicas, while the region B is propagated for a time β independetly for each replica (see Fig. 1.6). The factor Z^2 is just the partition function of two non-interacting replicas. Now it is clear that $S^{(2)}(\rho_A) = -\log(Z_A^{(2)}/Z^2)$ is proportional to the logarithm of the ratio of two partition functions (essentially a free energy difference) on the system of two replicas, which differ by the boundary conditions imposed on the region A in imaginary time.

We note that this replica-method approach for the computation of the Rényi entropies was first used in [41] in the context of conformal field theories.

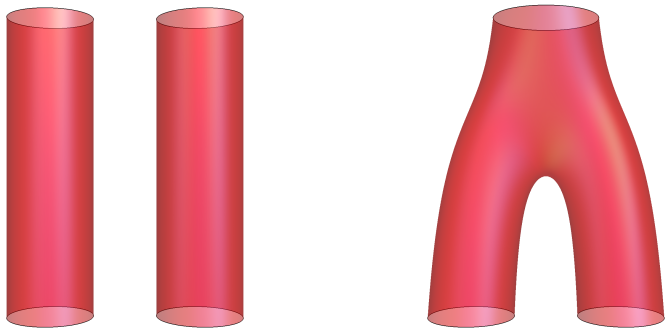


Figure 1.6: (Left) Pictorial impression of the path-integral representation of the Z^2 partition function of two replicas of a fictitious one-dimensional system. The vertical direction corresponds to the real-space coordinate x while the angular coordinate represents imaginary time τ. The imaginary-time coordinate is compact with length β for each replica. (Right) Pictorial representation of the system described by $Z_A^{(2)}$. Two copies of the original system are connected in the imaginary-time direction, but only in the region A (upper part of the surface), where the angular coordinate τ takes values in $0 \leq \tau < 2\beta$. The regions B (lower part of the surface) of the replicas are disconnected from each other, and the imaginary time loops close on themselves after an interval of length β. This gives a "pair of pants" Riemann surface representation of the system.

Chapter 2

Classical Ising Spin Glass

In this chapter we consider the classical Ising spin glass model defined on the Bethe lattice. This represent our theoretical approximation to the native (and only) problem that the D-Wave machine is able to solve. While this might seem to be a strong restriction for the computational capabilities of the machine, we show that this is, in fact, not the case. In Section 2.1 we review some classical models of spin glasses on generic interaction graphs. The infinite Bethe lattice, and the way to study it using finite-size regular random graphs as approximations, in presented in Section 2.1.1. The cavity-method solution of the model is discussed in Section 2.1.5 and the relation between Ising spin glasses and computational complexity is presented in Section 2.1.6.

2.1 The Ising spin glass problem

Consider a system of N classical spins, interacting according to the Hamiltonian

$$H = -\sum_{\langle ij \rangle} J_{ij} S_i S_j, \tag{2.1}$$

where $S_k = \pm 1$ are the spin variables and J_{ij} are quenched disordered variables sampled independently according to some fixed probability distribution $P(J_{ij})$. We assume that these spins interact only between nearest neighbours on a graph G, so that in Eq. (2.1) the sum index $\langle ij \rangle$ is taken over the edges of G. If for the disordered interaction couplings J_{ij} we choose the distribution

$$P(J_{ij}) = \frac{1}{2}\delta(J - J_{ij}) + \frac{1}{2}\delta(J + J_{ij})$$

for some fixed $J > 0$, then the model is called an Ising spin glass. As the Hamiltonian in Eq. (2.1) depends on the specific realization of disorder $J = \{J_{ij}\}$, then one is usually interested in computing some *disorder-averaged quantities* $\overline{A}$. These are obtained first by computing the value $A[J]$ for specific disorder realizations J, and then averaging this value with respect to the distribution of the disorder $\overline{A} \equiv \int A[J]P(J)\,\mathrm{d}J$. Quantities $A[J]$ whose variance over J disappears as one approaches the thermodynamic limit are called *self-averaging*.

Spin glasses are simplified models of disordered magnetic materials. The original paper [42] by Edwards and Anderson uses the model of Eq. (2.1) to

describe dilute magnetic alloys, where few impurities are assumed to be scattered throughout the crystal lattice structure of a solid. The magnetic dipoles of the atoms are then coupled by a potential that fluctuates rapidly with the separation between the atoms. If we model the dipoles as classical spin variables $S_i = \pm 1$ and the interaction between dipoles by random effective coupling constants J_{ij} with different signs (to represent the fluctuating potential) we obtain the Hamiltonian of Eq. (2.1).

Spin glasses are *frustrated systems*. A system is said to be frustrated if none of its states can simultaneously minimize the energy of every term in the Hamiltonian. By contrast, the ferromagnetic Ising model obtained from Eq. (2.1) by taking all couplings $J_{ij} = 1$ is not frustrated, as the ferromagnetically ordered configuration $S_i = 1$ for all i minimizes the energy of all terms in the sum. In the case where all terms give the same energy contribution, a ground state of a frustrated system is a spin configuration that minimizes the greatest possible number of interaction terms, and as such it is not guarateed to be unique. Moreover, one cannot expect to connect a given state to the ground state by following a local "steepest descent" approach, *i.e.* flipping one spin at a time and moving to the new state if its energy is lower than the energy of the current state. Indeed, if this method is applied one usually ends up in a *metastable state*, a state that is not the ground state but is sorrounded by states of higher energy (it is a local minimum of the energy).

A consequence of frustration, and one of the distinctive features of glassy systems, is the fact that they have complicated (or "rough") energy landscapes, in the sense that the energy suface defined over their configuration manifold has many hills and valleys, and therefore lots of local minima. In systems like the one in Eq. (2.1) where the configuration space is composed of a discrete set of points, this is meant to signify there are multiple ground states and many states with an energy that may be close, but not equal, to the ground state energy. These states can be separated (*e.g.* with respect to the Hamming distance) from the actual ground states by states of higher energy, so that these metastable states can function as "traps": if the system reaches one of these states then it will take a long time for the dynamics to move away from it[1].

One of the effects of these rough energy landscapes is that at temperatures lower than some critical temperature T_c, the free energy of the system develops a large number of valleys separated by barriers whose height diverges in the thermodynamic limit. This means that the phase space (or configuration space) is split into many regions that the dynamics cannot connect, *i.e.* the system becomes non-ergodic. This statement, however, must be taken with care, especially if one cannot work directly in the thermodynamic limit. This is because at any finite value of N, the system *is* ergodic, in the sense that, for an observable A, the long time-average

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t A(t')\,\mathrm{d}t'$$

[1] Of course, spin glass systems like the one in Eq. (2.1) do not have dynamics so the usual solution is to add an effective dynamics that is expected to model the microscopic dynamics of a system in a thermal environment, at least in the medium to large time regime. For spin systems, Metropolis or Glauber dynamics are commonly used.

is thermal, so the limit

$$\lim_{N\to\infty}\lim_{t\to\infty}\frac{1}{t}\int_0^t A(t')\,\mathrm{d}t'$$

is also thermal. The correct way of understanding glassy physics requires taking the limits in the opposite order, so that the integral is then

$$\lim_{t\to\infty}\lim_{N\to\infty}\frac{1}{t}\int_0^t A(t')\,\mathrm{d}t'.$$

In the previous equation, the $N\to\infty$ limit is evaluated first, which creates the clustering of the phase space into dynamically-separated regions. After this is done, the system is trapped in the same (initial-state dependent) region for all times t, and in the $t\to\infty$ limit the system mixes (*i.e.* "thermalizes") only within the states that lie inside of this region. The time-averages then converge to an average with respect to this reduced ensemble (sometimes called a "pure state"), rather than the usual canonical ensemble.

While the dynamical point of view is the most traditional, glassiness can be understood from the perspective of equilibrium properties. Again, it is useful to compare their behaviour with the behaviour of ferromagnatic systems, described by the Hamitonian (2.1) but where all interactions are positive $J_{ij}=1$. If $\langle\cdot\rangle_\beta$ is the canonical-ensemble average at inverse temperature β, then the magnetization is commonly defined as

$$m\equiv\frac{1}{N}\sum_i m_i=\frac{1}{N}\sum_i\langle S_i\rangle_\beta$$

The Hamiltonian of (2.1) is $\mathbb{Z}_2$-symmetric with respect to the global spin-flip operation that maps $S_i\to -S_i$ simultaneously for all sites i. This means in particular that $\langle S_i\rangle_\beta=0$ for all i, so the magnetization is always zero, at all temperatures. This of course fails to capture the well-established experimental fact that ferromagnets exhibit spontaneous magnetization at low temperatures. The usual solution is then to apply a small magnetic field $H_{\text{long}}=-h\sum_i S_i$ to the system, where $h\in\mathbb{R}$, and then bring it to zero. Then for a choice of sign $s=\pm$ one defines

$$m_i(s)\equiv\lim_{h\to 0^s}\lim_{N\to\infty}\langle S_i\rangle_{\beta,h}$$

where by $\langle\cdot\rangle_{\beta,h}$ is meant the Gibbs distribution at inverse temperature β induced by the Hamiltonian of Eq. (2.1) with the additional term H_{long} multiplied by the chosen sign s. What happens is that at high temperatures, all the local magnetizations m_i are zero. At low temperatures, however, the local magnetizations will all assume the same finite value $m_i=\pm m(\beta)$ whose sign is selected by the sign of the external field h. This is because as the temperature is lowered below the Curie critical temperature T_c, the Gibbs distribution will concentrate on the low-energy states, which in the case of ferromagnets have large domains of aligned spins. This means that the spins will all tend to align in the same direction, *e.g.* $\langle S_i\rangle=1$, and the total magnetization will assume a finite value. Therefore, in the physics of ferromagnets, the phase transition between a paramagnetic and an ordered phase is signalled by the value of total magnetization $m=(1/N)\sum_i m_i$. The emergence of a finite value for the magnetization signifies the beginning of the ordered phase.

The same temperature-driven spin alignment effect also happens in the case of spin glasses. However, the low-energy states of a spin glass usually does not have the large domains of aligned spins that are typical of ferromagnetic systems. Instead, the spins are aligned in random directions, so that in the limit $\beta \to \infty$ the average local magnetization $\langle S_i \rangle$ at site i will approach essentially a random value of $+1$ or -1, independently for each site i. This results in a zero total magnetization even at low temperatures, so it cannot be used as an order parameter for the transition. Instead, one uses the *Edwards-Anderson order parameter* q_{EA}

$$q_{EA} \equiv \frac{1}{N} \sum_i \langle S_i \rangle^2 .$$

Again, this is to be intended as the value of a limiting procedure where the breaking of ergodicity is introduced in some way. Either in the static way, where one introduces local external fields to the Hamiltonian, $H_{long} = -h \sum_i h_i S_i$ for a fixed choice of $\vec{h} = (h_1, \ldots, h_N) \in \{0, 1\}^N$ and $h \in \mathbb{R}$, and then takes the limit of vanishing field strength

$$m_i(\vec{h}) \equiv \lim_{h \to 0^+} \lim_{N \to \infty} \langle S_i \rangle_{\beta, \vec{h}}$$

so that $q_{EA}(\vec{h}) = (1/N) \sum_i m_i(\vec{h})^2$, or in the dynamical approach

$$q_{EA} = \lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{N} \sum_i \left(\frac{1}{t} \int_0^t S_i(t') \, \mathrm{d}t' \right)^2 .$$

A *spin glass phase* is defined as a temperature interval where q_{EA} is non-zero.

While interesting from a purely physical perspective, our interest in spin glasses stems also from the relation they have with combinatorial optimization problems. On the one hand, it is easy to see that (2.1) can be recast as constraint satisfaction problem of Section 2.1. An interaction term $-J_{ij} S_i S_j$ becomes a constraint f_{ij} over the two Boolean variables $X_i \equiv (1 - S_i)/2$ and $X_j \equiv (1 - S_j)/2$ such that

$$f_{ij}(X_i, X_j) = \begin{cases} -J_{ij} & \text{if } X_i = X_j \\ J_{ij} & \text{otherwise} \end{cases} ,$$

and the full Hamiltonian is given by $f(X_1, \ldots, X_N) = \sum_{\langle i,j \rangle} f_{ij}(X_i, X_j)$. Then solving this minimization problem is equivalent to finding the ground state of the corresponding spin glass Hamiltonian.

On the other hand, and perhaps more surprisingly, a great number of optimization problems can be reformulated as instances of a spin glass problem, to the effect that one needs only to be able to solve the spin glass case to be able to solve all of these other problems. This is the reason why the D-Wave machine – which can solve only this energy minimization problem – is more than a curiosity. This will be better explained in Section 2.1.6.

2.1.1 The Bethe lattice and Regular Random Graphs

Since the spin glass model we study in this thesis is defined on graphs with non-trivial topologies, let us quickly review some relevant graph-theoretical definitions. An *undirected graph* (from now on, simply "a graph") $G = (V, E)$ is

defined by a set of vertices V and a set of edges E, *i.e.* subsets of V with exactly two elements. If $e = \{v, w\}$ is an edge of G, then we say that the vertices v and w are *adiacent* (or *neighbours*), and that the edge e is *incident* to both v and w. The *degree* of a vertex, $\deg(v)$ is the number of vertices it is adiacent to. If all vertices of G have the same degree d, then G is said to be *regular* with $\deg(G) = d$. A finite graph $G = (V, E)$ over the vertices $V = \{v_1, \dots, v_N\}$ can be represented by an operator A_G over $\mathbb{C}^N$ called the *adjacency matrix* and defined by

$$\langle v_i | A_G | v_j \rangle = \begin{cases} 1 & \text{if } \{v_i, v_j\} \text{ is an edge in } E \\ 0 & \text{otherwise} \end{cases},$$

where $\{|v_i\rangle\}_{i=1}^N$ is any fixed basis whose elements are labelled by the vertices v_i.

Given a graph $G = (V, E)$ and two vertices $v, w \in V$, a "path" p connecting v and w is a sequence of vertices $p = (v_0, v_1, \dots, v_n)$ such that $v_0 = v, v_n = w$ and for every $0 \leq i \leq n-1$ it holds that v_i and v_{i+1} are adiacent in G. Alternatively, the same path p can be defined by the sequence of edges $p = (e_0, \dots, e_{n-1})$ it crosses, where $e_i \equiv v_i, v_{i+1}$. The number n of edges in a path p is defined to be the *length* of the path and denoted $\ell(p)$. We write $\text{paths}(v, w)$ to indicate the set of paths connecting v and w. A path starting and ending in the same vertex $v = v_0 = v_n$ is called a *cycle*, or *loop*. A graph without cycles is called a *tree*.

A graph is *connected* if every pair of its vertices has a path that connects them, *i.e.* $\text{paths}(v, w)$ is non-empty for all v, w. Connected graphs have automatically a metric structure induced by the path lengths. Given two vertices v, w one defines

$$\text{dist}_G(v, w) \equiv \min \Big\{ \ell(p) \mid p \in \text{paths}(v, w) \Big\}.$$

This can easily seen to satisfy the usual conditions for a distance function. The open balls for this metric space are defined as usual

$$B(v, r) \equiv \{w \in V \mid \text{dist}_G(v, w) < r\},$$

where $v \in V$ is the center and r is the radius.

In this thesis we want to consider the spin glass problem of Eq. (2.1) where sum is taken over the edges of a Bethe lattice. The Bethe lattice (see Fig. 1.1)is a graph that was introduced in [43] to define models where the Bethe-Peierls approximation is exact.

Definition 1 (Bethe lattice)**.** *The Bethe lattice $\mathbb{T}_d$ of degree d is a connected tree with a countably infinite set of vertices, with the property that each vertex has exactly d neighbours (i.e. is d-regular). The subgraph obtained from $\mathbb{T}_d$ by removing any single vertex is a collection of d isomorphic finitary trees. The branching factor of these trees, usually called the* connectivity *of the Bethe lattice in the physics literature and indicated by the letter K, is equal to $d - 1$.*

Note that the Bethe lattice is locally homogeneous, in the sense that any pair of balls $B(v_1, r), B(v_2, r)$ of the same radius r centered on vertices v_1, v_2, are isomorphic. While its highly symmetric structure is a boon for analytical methods, the fact that it is infinite in size makes it difficult to study numerically. Therefore one would like to study a sequence $\{G_N\}_{N \in \mathbb{N}}$ of graphs of finite size N that in some sense converge to the Bethe lattice $\mathbb{T}_d$ in the limit $N \to \infty$.

In this way, the value of some physical quantity of interest Q can be computed for the graphs in the sequence, obtaining the values $Q_1, Q_2, Q_3, \ldots$. The limit $Q_\infty \equiv \lim_{N\to\infty} Q_N$ is then the value of the quantity Q computed on the inifinite-size Bethe lattice $\mathbb{T}_d$[2]. For the case of disordered systems, the usual choice is to use *regular random graphs* of finite size N and fixed degree d.

Definition 2 (Random Regular Graph)**.** *For fixed integers $k, N > 0$, let $\mathcal{G}_{d,N}$ be the set of all graphs with N vertices so that each vertex has exactly d neighbours. Note that for any choice of k, N, the set $\mathcal{G}_{d,N}$ is finite, so we can equip it with a uniform probability distribution*

$$P_{d,N}(G) = \frac{1}{|\mathcal{G}_{d,N}|} \qquad \text{for any } G \in \mathcal{G}_{d,N},$$

and the complete sigma-algebra $\Sigma \equiv \mathcal{P}(\mathcal{G}_{d,N})$ so that $(\mathcal{G}_{d,N}, \Sigma, P_{d,N})$ is a probability space. A "d-regular random graph of size N" is any graph G sampled from this probability space.

2.1.2 Locally tree-like graphs

Unlike the Bethe lattice, regular random graphs can be *loopy*, which means that one can find a path $(v_1, \ldots, v_k)$ starting and ending on the same vertex $v_1 = v_k$. The presence of loops in an interaction graph has profound consequences for the physics of disodered systems, as this is a way of introducing frustration. This is shown in Fig. 2.1 with the simplest example of frustrated system is given by a triangle (a loop of length 3) with antiferromagnetic interactions.

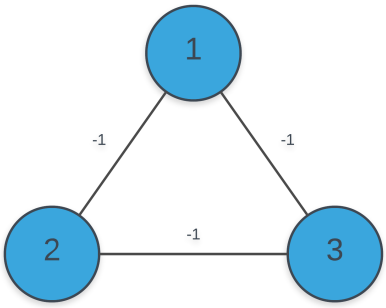


Figure 2.1: one of the simplest frustrated system, the triangle with antiferromagnetic interactions. The three terms in the Hamiltonian $H = S_1S_2 + S_2S_3 + S_1S_3$ cannot be minimized simultaneously. This is a general behaviour that happens frequently in antiferromagnetic and disordered systems, and gives rise to rough energy landscapes.

One of the properties of large random regular graphs is that, while loopy, they are locally tree-like: given any fixed radius r and a vertex v in any large

[2] in the physics literature, the value of this limit is *defined* to be the value of Q on Bethe lattice even in the case where the physical system of interest is ill-defined on an actually-infinite graph.

enough RRG, the loops in the graph will, with high probability, concentrate outside of the ball of radius r centered in v. This means that the presence of a loop in the graph cannot be detected by looking at any finite-radius neighbourhood of a vertex, because (in the $N \to \infty$ limit) any such local neighbourhood will be a tree. Loops will only show up if one looks at neighborhoods whose radius r grows with N. More rigorously, one can prove (see Corollary 2 in [44]) the following theorem.

Theorem 1. *Let $d, m \geq 3$ and $\ell \geq 0$, and let $\mathrm{Cycles}(G, \ell)$ be the set of loops of length ℓ in a graph G. Then in the $N \to \infty$ limit, the probability that a random graph $G \in \mathcal{G}_{d,N}$ has exactly k cycles of length ℓ is given by*

$$\lim_{N\to\infty} \Pr_{\mathcal{G}_{d,N}} \Big[|\mathrm{Cycles}(G,\ell)| = k \Big] = \frac{\lambda_\ell^k e^{-\lambda_\ell}}{k!},$$

where $\lambda_\ell \equiv (d-1)^\ell/(2\ell)$.

Moreover, the random variables $X_\ell \equiv \mathrm{Cycles}(\cdot, \ell)$ are statistically independent over $\mathcal{G}_{d,N}$ with $\langle X_\ell \rangle = \lambda_\ell$. Since the expectation value of the number of cycles of a fixed length ℓ is finite for $N \to \infty$, and the number of cycles in G almost-surely diverges with N, then the probability that a random cycle in G has length ℓ tends to zero in the $N \to \infty$.

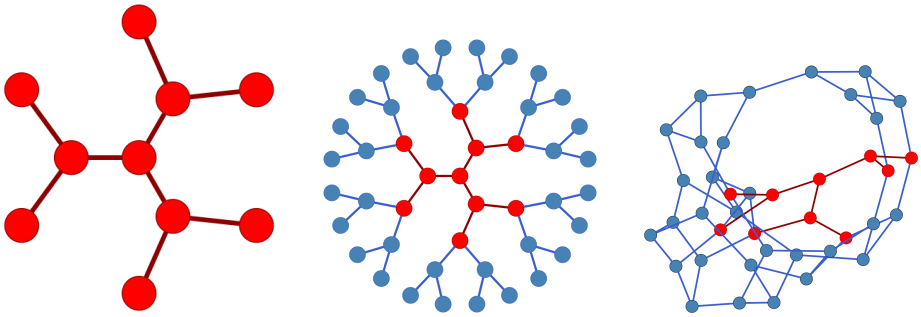


Figure 2.2: Locally, both the Bethe lattice and a regular random graph have a tree structure. The loops that are present in a regular random graphs appear only at the global scale.

2.1.3 Benjamini-Schramm convergence

For a given graph G, the two conditions of 1) having a fixed degree, and 2) being locally tree-like suggest that G should locally look like the Bethe lattice of the same degree. We have seen that both conditions hold, with high probability, for large regular random graphs. One gets the intuition that a typical regular random graph should share the same properties of the Bethe lattice, at least as long as local properties are considered. This intuitive idea can be formalized using a limiting procedure for graph sequences called Benjamini-Schramm (BS) convergence [45, 46].

In order to define BS-convergence one has to use the somewhat technical notion of a rooted graph. A *rooted graph* is a pair (G, r) where $G = (V, E)$ is a

graph and $r \in V$ is a distinguished vertex called the *root*. An *isomorphism of rooted graphs* $\phi : (G_1, r_1) \to (G_2, r_2)$ is a graph isomorphism where $\phi(r_1) = r_2$. We write $\mathcal{RG}_d$ to indicate the set of all d-regular rooted graphs of arbitrary size. This set can be given a metric structure (and therefore a topological structure) by defining a distance function $d(G_1, G_2)$ based on the distance one needs to moves away from the respective roots of G_1 and G_2 before the two graphs can be recognized to be non-isomorphic.

Definition 3. *Given two rooted graphs* (G_1, r_1) *and* (G_2, r_2)*, consider the set*

$$\mathcal{S}(G_1, G_2) = \{k \in \mathbb{N} \mid B(r_1, k) \cong B(r_2, k)\}$$

of values $k \in \mathbb{N}$ *such that the balls* $B(r_1, k)$ *centered on the root of* G_1 *and* $B(r_2, k)$ *centered on the root of* G_2*, are isomorphic. We define the distance function* $d(G_1, G_2) \equiv 1/k_{\max}$*, where* $k_{\max}$ *is defined as*

$$k_{\max} \equiv \sup \mathcal{S}(G_1, G_2)$$

with the understanding that $d(G_1, G_2) = 0$ *if* $\mathcal{S}(G_1, G_2)$ *has no upper bound.*

One can prove that this is a *bona fide* distance function if one defines it over $\mathcal{RG}_d$ modulo graph isomorphisms. This makes $\mathcal{RG}_d$ (or rather $\mathcal{RG}_d/\cong$) a metric space.

The Benjamini-Schramm convergence is the weak convergence induced by the probability measures on the space $\mathcal{RG}_d$ of rooted graphs: we call *generalized random graph* of degree d a probability measure μ over $\mathcal{RG}_d$. Then the sequence of generalized random graphs $\{\mu_N\}_N$ converges in the BS sense to the generalized random graph μ if, for all continuous functions $f : \mathcal{RG} \to \mathbb{R}$ it holds that

$$\lim_{N\to\infty} \int_{\mathcal{RG}} f(g)\, \mathrm{d}\mu_N(g) = \int_{\mathcal{RG}} f(g)\, \mathrm{d}\mu(g).$$

Equivalently, one can use finite rooted graphs to check BS-convergence. Given a generalized random graph μ, a finite rooted graph (G, r) and an integer $k > 0$, define

$$P(\mu, G) \equiv \Pr_\mu \Big[B(r, k) \cong G\Big]$$

as the probability that a ball of radius k centered at the root r of a rooted graph $g \in \mathcal{RG}_d$ sampled according to μ, is isomorphic to the rooted graph (G, r). Then one can prove that the sequence $\{\mu_N\}_N$ converges in the BS sense to the generalized random graph μ if and only if for all finite rooted graph (G, r) and all $k > 0$ it holds that

$$\lim_{N\to\infty} P(\mu_N, G) = P(\mu, G).$$

Then by taking μ_N to be the uniform distribution over d-regular random graphs of size N, with randomly chosen roots, and $\mu = \mathbb{T}_d$ (*i.e.* a delta function centered on the d-regular Bethe lattice with any fixed choice of root), then one can see that Theorem 1 implies that (randomly rooted) d-regular random graphs converge to the Bethe lattice $\mathbb{T}_d$ with respect to the Benjamini-Schramm convergence.

In light of the previous results, we will use regular random graphs as finite-size approximations of the Bethe lattice, and we will claim that the thermodynamic-limit value of some quantity computed on the finite-size RRG ensembles $\{\mathcal{G}_{d,N}\}$

applies to the Bethe lattice as well. However, one must take care when doing this as the Benjamini-Schramm convergence preserves only *local properties*. Which exactly are to be considered local properties is an open question in combinatorics. From a physical point of view, one can argue that bulk properties of physical systems described by local Hamiltonians such as that of Eq. (2.1) are conserved by taking the Benjamini-Schramm limit.

2.1.4 Expander graphs

Random regular graphs have one more property that will be relevant to us, namely they are well-connected in a sense that can be formalized using the graph-theoretic equivalent of the Cheeger isoperimetric constant for Riemann manifolds. Let $\mathcal{M}$ be a closed Riemann manifold with $\dim(\mathcal{M}) = n$. For a given submanifold $X \subseteq \mathcal{M}$, let $Vol(X)$ be the volume of X:

$$Vol(X) = \int_{\mathcal{M}} \chi_X \mathrm{d}V$$

and let $Area(X)$ be the $(n-1)$-dimensional volume of the boundary ∂X of the submanifold X induced by the volume form $\mathrm{d}V$. Then one defines

$$h_{\mathcal{M}}(X) \equiv \frac{Area(X)}{\min\left\{Vol(X), Vol(X^c)\right\}}$$

where X^c is the complementary closed manifold to X in $\mathcal{M}$. Then the Cheeger constant $h_{\mathcal{M}}$ of the manifold $\mathcal{M}$ is defined as

$$h_{\mathcal{M}} \equiv \inf_X h_{\mathcal{M}}(X)$$

where the infimum is taken over all closed submanifolds of $\mathcal{M}$. In the case of graphs, a similar quantity can be defined.

Definition 4 (Graph-theoretic Cheeger constant)**.** *Let $G = (V, E)$ a graph and, for any $A, B \subseteq V$, define $E(A, B)$ as set of edges of G incident to a vertex in A and a vertex in B. Given a subset of vertices $W \subseteq V$, define the quantity*

$$h_G(W) \equiv \frac{|E(W, W^c)|}{\min\left\{\sum_{v\in W} \deg(v), \sum_{v\in W^c} \deg(v)\right\}}$$

where W^c denotes the complement of the set W in V. Then the "Cheeger constant" of the graph G is defined as

$$h_G \equiv \min_{W\subseteq V} h_G(W)$$

This is equivalent to the Riemann manifold definition, with the prescription that $E(X, X^c)$ measures the area of the boundary of a set of vertices X and that $\sum_{x\in X} \deg(x)$ measures the volume of X. Note that disconnected graphs have $h_G = 0$. Intuitively, graphs with a small positive value of h_G have a "bottleneck", *i.e.* a pair of regions connected only by relatively few edges. On the opposite regime, a large h_G means that every two regions of the graph are connected by relatively many edges. Families of graphs $\{G_n\}$ with a Cheeger constant that is bounded away from zero are called *expander graphs*. The main point then is that large regular random graph are (almost surely) good expanders, as summarized in the following result, adapted from [47]

Theorem 2. *Let $d \geq 3$ and let $G \in \mathcal{G}_{d,N}$ a random regular graph. Then the modulus largest non-trivial eigenvalue $\lambda(G)$ of the adiacency matrix of G (i.e. $\lambda(G) \neq d$) satisfies*

$$\lambda(G) < \sqrt{(d-1)} + 1$$

asymptotically almost surely.

One can then show that this bound on $\lambda(G)$ implies a bound on the Cheeger constant $h(G)$, as it is well known that spectral expansion is related to edge expansion (see *e.g.* [48]).

2.1.5 Ising Spin Glass on the Bethe Lattice

The Ising spin glass on the Bethe lattice $\mathbb{T}_d$ enters a glassy phase at low temperatures, when $\beta > \beta_c$, where the critical value of the transition is given by the formula

$$\tanh(\beta_c) = 1/\sqrt{K}, \tag{2.2}$$

and $K = d - 1$ is the connectivity of the lattice. First of all, we recall that the system can be solved numerically, directly in the thermodynamic limit, to arbitrary degrees of accuracy with the cavity method. While the solution requires the breaking of the replica symmetry (see *e.g.* [49]), the simpler replica-symmetric approach is enough to show the emergence of the spin glass phase, which will be enough for us.

In the simplest, replica-symmetric version of the cavity method, one considers the marginal of the Gibbs distribution on a given site, say $i = 0$

$$P(S_0) = \frac{1}{Z} \sum_{S_k : k \neq 0} e^{\beta \sum_{\langle ij \rangle} J_{ij} S_i S_j}$$

and reparametrizes it with an *effective field* (also called *cavity field*) parameter h_0 such that

$$P(S_0) = e^{\beta h_0}.$$

It is possible to find a recursion equation for the cavity fields of the spin 0 once those of the adiacent spins $1, \ldots, K$ are known

$$h_0 = \frac{1}{\beta} \sum_{i=1}^{K} \tanh^{-1}(\tanh(\beta J_{0i}) \tanh(\beta h_i)). \tag{2.3}$$

These fields are random i.i.d. (since the correlations are negligible for large system sizes) and distributed according to a probability distribution which is stable under (2.3). The distribution which is stable under iterations in the paramagnetic phase is

$$P(h) = \delta(h), \tag{2.4}$$

and the limit of stability of this solution is observed by expanding Eq. (2.3) for small h_i

$$h_0 \simeq \sum_{i=1}^{K} \tanh(\beta J_{0i}) h_i. \tag{2.5}$$

By squaring and taking the average both over J and over $P(h)$ we can see how the size of the distribution (as measured by $\langle h^2 \rangle$, given that $\langle h \rangle = 0$) evolves under the iterations. We get

$$\langle h^2 \rangle' = K \tanh^2(\beta J) \langle h^2 \rangle \tag{2.6}$$

which means that until $T > T_c$ given by $T_c \equiv 1/\beta_c$ in Eq. (2.2), the stable distribution has a decreasing $\langle h^2 \rangle$ until $\langle h^2 \rangle \to 0$, from which we obtain ((2.4)) [50]. Proceeding with this reasoning one can also find that for $T \lesssim T_c$:

$$q_{EA} \equiv \frac{1}{N} \sum_i \langle S_i \rangle^2 \propto |T_c - T|. \tag{2.7}$$

Another instructive way to obtain the same results is to consider the spin-spin connected correlation function

$$C_{ij} \equiv \langle \sigma_i^z \sigma_j^z \rangle - \langle \sigma_i^z \rangle \langle \sigma_j^z \rangle = \frac{1}{\beta} \frac{\partial \langle \sigma_i^z \rangle}{\partial \eta_j}, \tag{2.8}$$

where η_j is a local field applied on the spin j (and then sent to zero) $H \to H - \eta_j \sigma_j^z$. By using a telescopic identity and the recursion relations on the field, for two spins i, j separated by a path of vertices $\{v_n\}_{n=0}^L$ of length $L \equiv d(i,j)$:

$$j \equiv v_0, v_1, \ldots, v_L \equiv i$$

we have (denoting $v_n \equiv j + n$ and $v_{L-m} \equiv i - m$) that the main contribution to the correlation function is given by

$$C_{ij} = \beta^{-1} \frac{\partial \langle \sigma_i^z \rangle}{\partial h_{i-1}} \frac{\partial h_j}{\partial \eta_j} \prod_{k=1}^{L} \frac{\partial h_{j+k}}{\partial h_{j+k-1}}. \tag{2.9}$$

Now

$$\frac{\partial h_a}{\partial h_{a-1}} = \frac{\tanh(\beta J_{a,a-1}) \cosh^{-2}(\beta h_{a-1})}{1 - \tanh(\beta J_{a,a-1})^2 \tanh^2(\beta h_{a-1})}. \tag{2.10}$$

Since we are studying the stability of the paramagnetic phase and the critical region we can set $h_a \to 0$, which gives

$$\frac{\partial h_a}{\partial h_{a-1}} \simeq \tanh(\beta J_{a,a-1}). \tag{2.11}$$

Considering that at a distance $L \sim \ln N / \ln K \gg 1$, diameter of the RRG, there are K^L paths that lead from one spin to another, we have that the susceptibility on a single path has to be summed over all the K^L paths p

$$C_{ij} = \beta^{-1} \frac{\partial \langle \sigma_i^z \rangle}{\partial \eta_j} \propto \sum_{p=1}^{K^L} \prod_{a_p=1}^{L} \tanh(\beta J_{a_p, a_p - 1}). \tag{2.12}$$

Since this is a sum of randomly signed, i.i.d. terms, we have that the typical value is

$$\sqrt{\left\langle C_{i,i+L}^2 \right\rangle} \sim K^{L/2} \tanh(\beta J)^L. \tag{2.13}$$

This decays exponentially if and only if $\beta < \beta_c$ in Eq. (2.2). Notice that this sum is not dominated by a single term, but it is a collective behavior of the single terms which gives rise to the transition.
This correlation function can be used to define the *shattered susceptibility*

$$\chi_s = \frac{1}{N} \sum_{ij} C_{ij}^2, \tag{2.14}$$

which diverges at the transition and remains infinite in the whole SG phase, below the AT line [51, 50]. On the Bethe lattice this is due to the fact that the exponential growth of the number of sites at distance r, K^r, dominates over the exponential decay of the typical spin-spin correlation.

2.1.6 Computational Complexity

From a computational perspective, the spin model of Eq. (2.1) is interesting because finding or approximating its ground state energy is a hard optimization problem in the worst-case. We have remarked before that the D-Wave machine's only purpose is to find the ground state of a classical spin glass Hamiltonian. Even though this could be hard to do, it would seem *prima facie* to be a problem with rather limited applications. Actually, the opposite is true, as it turns out that solving this problem efficiently would provide us a way of solving a vast set of problems called NP.

Informally, NP is the class of decision problems whose candidates for solutions are easy to check, but whose solutions might be hard to find. As an example, take integer factorization: suppose we are given two integers N and M and we are asked whether N has a divisor K such that $1 < K < M$. Deciding this could be hard as one would need to find such a divisor among many possible candidates. However, if we are given an integer K as a candidate solution for the problem and asked whether it is indeed a solution, then this is much easier as one can simply use the Euclidian algorithm to check if K divides N (also, the condition $1 < K < M$ is trivial to check).

NP-hard problems are in a sense at least as hard as any problem in NP: a problem P is NP-hard if for all problems $Q \in \mathsf{NP}$ there is a way to map instances of Q to instances P so that if one can efficiently solve instances of the problem P then they are also able to solve efficiently the corresponding instances of Q. As NP is a very large class of problems, being able to solve efficiently even *one* NP-hard problem would be an extremely significant result.

The NP-hardness of the Ising spin glass problem follows from the fact that it is dual to the well-known NP-hard problem Weighted Max-Cut with weights $w = \pm J$, in the sense that we now explain.

Given a graph $G = (V, E)$, a *cut* is a subset $W \subseteq V$ of the vertices of G, and its size is defined to be the number of edges incident to one vertex in W and one vertex in its complement $W^c = V \setminus W$. In the notation of Subsection 2.1.4

$$\text{size}(W) = |E(W, W^c)|.$$

The Max-Cut problem is to find a cut with the largest size. If the edges $e \in E$ of the graph G are assigned a weight $w(e) \in \mathbb{R}$, then one defines the size of a

cut by the sum of the weight of its boundary:

$$\text{size}(W) = \sum_{e \in E(W,W^c)} w(e).$$

Then the WEIGHTED MAX-CUT problem is to find a cut with the largest size, according to this second meaning. The WEIGHTED MAX-CUT can be cast as a decision problem (admitting a yes/no answer) in the usual way.

> WEIGHTD MAX-CUT
>
> **Instance:** a graph $G = (V, E)$ with weighted edges $w(e) \in \mathbb{R}$. A real number x.
> **Question:** is there a cut $W \subseteq V$ such that $\sum_{e \in E(W,W^c)} w(e)$ is larger than x?

Given a specific realization of Eq. (2.1) defined on a generic graph $G = (V, E)$ and the energy of a specific spin configuration $\vec{S}$

$$H(\vec{S}) = -\sum_{\langle i,j \rangle} J_{ij} S_i S_j,$$

one can split the sum into the edges that connect spins that are aligned in the configuration $\vec{S}$, and the edges that connect anti-aligned spins:

$$H(\vec{S}) = -\sum_{\langle i,j \rangle : S_i = S_j} J_{ij} + \sum_{\langle i,j \rangle : S_i \neq S_j} J_{ij}.$$

Now, we can add the same configuration-independent quantity $\sum_{\langle i,j \rangle} J_{ij}$ to both sides of the equation

$$\begin{aligned} H(\vec{S}) + \sum_{\langle i,j \rangle} J_{ij} &= -\sum_{\langle i,j \rangle : S_i = S_j} J_{ij} + \sum_{\langle i,j \rangle : S_i \neq S_j} J_{ij} + \sum_{\langle i,j \rangle} J_{ij} \\ &= 2 \sum_{\langle i,j \rangle : S_i \neq S_j} J_{ij}. \end{aligned}$$

So if we associate a spin configuration $\vec{S}$ to a set of vertices $X_{\vec{S}} \subseteq V$

$$\vec{S} \leftrightarrow \{i \in V \mid S_i = +1\} \equiv X_{\vec{S}}$$

and define

$$W(X_{\vec{S}}) \equiv \frac{1}{2}\Big(H(\vec{S}) + \sum_{\langle i,j \rangle} J_{ij}\Big) = \sum_{\langle i,j \rangle : S_i \neq S_j} J_{ij}, \tag{2.15}$$

we get a WEIGHTED MAX-CUT instance with weighted edges $w(i, j) = J_{ij}$. The affine relation (2.15) between energy $H(\vec{S})$ and weight $W(X_{\vec{S}})$ entails in particular that

$$\arg\min_S H(\vec{S}) = \arg\max_S W(X_{\vec{S}}).$$

Moreover, since WEIGHTED MAX-CUT is NP-hard this is also true for the following decision problem.

ISING SPIN GLASS

Instance: a graph $G = (V, E)$ with weighted edges $J_{ij} = \pm J$, with $J > 0$ a real number. A real number E.
Question: is there a spin configuration $\vec{S} \in \{\pm 1\}^V$ such that $H(\vec{S}) = -\sum_{\langle i,j \rangle} J_{ij} S_i S_j$ is smaller than E?

Therefore finding the ground state energy of an Ising spin glass system defined on a generic interaction graph is NP-hard, as this would allow one to solve the ISING SPIN GLASS decision problem. For comparison, finding the ground state energy of the same Ising spin glass Hamiltonian defined on planar or toroidal graphs is a problem that can be solved exactly in polynomial time [52] (that is, easily).

Even when exact solutions are unknown or unlikely to exist one can often find some approximating algorithm that gives a solution close to the exact one. Sometimes one can even find arbitrarily good approximations, such as in the case of a N-vertex cubic lattice in three dimensions, whose ground-state energy is $O(\epsilon N)$-approximable for any $\epsilon > 0$. In a nutshell, one can partition the lattice into small cubes of size $L \times L \times L$ and define a new Hamiltonian H' by discarding from H the interaction terms that cross from one cube into another. Then H' is a sum of terms defined on different, non-interacting regions of the system and one can solve each term separately, even by brute force (e.g. checking the energy of each possible configuration and taking one with minimal energy). This takes linear time since there are $O(N)$ cubes and each one requires 2^{L^3} steps which is a constant in N. Finally one gives the ground-state energy of H' as the approximation of the ground-state energy of the original Hamiltonian H. The number of bonds discarder by the approximation is given by

$$\frac{1}{2} \, |\partial_{cube}| \times N_{cubes} = 3L^2 \frac{N}{L^3} = \frac{3N}{L} = O\Big(\frac{N}{L}\Big)$$

where N_{cubes} is the number of small cubes and $|\partial_{cube}|$ is the number of bonds across the boundary of a single cube. Since $|J_{ij}| = J$ for all bonds (i, j) then the absolute error of the estimated ground-state energy is upper bounded by $3JN/L$ and one can then choose a large enough L so that the $O(N)$ prefactor is as small as desired.

The success of this approximation is based on the fact that surface effects can be neglected since these grow like L^2 while volume effects grow like L^3. However, this condition is not satisfied in the case of a low-degree regular random graph because large RRGs are with high probability good expander graphs [53], so the boundary of any region is proportional to the volume of that region. Physical systems whose interaction graphs are expanders have boundary effects that cannot be neglected. In the cubic lattice example, if we had that $|\partial_{cube}| \propto L^3$ then we would get an absolute error which is independent of L and therefore the previous approximation scheme cannot be fine-tuned to achieve any desired error. From a computational complexity perspective this is in fact to be expected, as WEIGHTED MAX-CUT is known to be APX-hard, meaning that (under the standard assumption that P $\neq$ NP) it does not have a

polynomial-time approximation scheme (PTAS) capable of approximating the optimal solution to arbitrarily small error.

Chapter 3

Thermodynamics of the Quantum Ising Spin Glass

In this chapter we focus on studying the equilibrium properties of the quantum version of the Ising spin glass on the Bethe lattice we described in Chapter 2. In Section 3.1 we quantize the classical Hamiltonian by introducing a transverse field and discuss the relevant literature of the model we obtain. This model transitions from a paramagnetic to a spin glass phase at small temperature and transverse field. Section 3.2 describes the physical quantities we are interested in computing in order to study the quantum phase transition between the paramagnetic and the glassy phase of this model, as well as the numerical methods we use to compute them. Section 3.3 describes a perturbation expansion we employ to explain the quantitative results obtained from the numerics and get a better understanding of the critical point.

3.1 Quantization of a classical spin glass

The usual way to quantize a spin system Hamiltonian such as (2.1) is to choose a quantization axis (in our case it will be the z-axis) and promote a classical spins S_i to a Pauli operator σ_i^z along this axis:

$$H = -\sum_{\langle ij \rangle} J_{ij} \sigma_i^z \sigma_j^z.$$

In order to give the system non-trivial dynamics we include an additional, non-commuting term to this Hamiltonian. The simplest example is a transverse-field term $-\Gamma \sum_i \sigma_i^x$. We obtain the Hamiltonian

$$H = -\sum_{\langle i,j \rangle} J_{ij} \sigma_i^z \sigma_j^z - \Gamma \sum_{i=1}^{N} \sigma_i^x, \tag{3.1}$$

where $\Gamma > 0$ is the strength of the transverse field, and σ_i^a (for $a = x, z$) is a Pauli matrix acting on the i-th spin of the system. We remind the reader that this is the typical example of a Hamiltonian used in a Quantum Annealing protocol, where $\Gamma = \Gamma(t)$ is large at the initial time $t = 0$ and is then slowly

decreased to zero. If we start in the ground state of the transverse-field term, then at the end of the process the system will be in the ground state of the spin glass term, which encodes the solution to the combinatorial problem we are trying to solve.

In this chapter we study the equilibrium properties of the quantized version of the Ising spin glass Hamiltonian we described in the previous chapter, *i.e.* a system of N interacting quantum spins with quenched disorder, affected by a transverse field, described by the Hamiltonian of Eq. (3.1). The disordered interaction couplings J_{ij} are independently, identically distributed variables that take either of the two values in $\{\pm 1\}$ with equal probability. The first sum index $\langle i, j\rangle$ is taken over the edges of an interaction graph that we take to be a regular random graph, which for convenience will have degree $d = 3$ in the rest of this thesis. This particular choice is dictated by the numerical treatments, since this allows us to have systems of larger "linear dimension", the diameter of the regular random graph $L \simeq \ln N / \ln K$.

This model is no longer exactly solvable but it was studied in [54] using a quantum version of the cavity method, where it was found to exhibit a transition from a paramagnetic to a glassy phase for small values of the transverse field Γ and temperature T. The method is exact on trees and is expected to be a good approximation on locally tree-like graphs. However, it becomes less reliable close to the $T = 0$ line and an estimate for the zero-temperature Γ_c was given as lying between 1.5 and 2.0. In [55] the antiferromagnetic version of Eq. (3.1) was studied, where all $J_{ij} = -1$. This is known as the 3-Max-Cut problem in the computer science community. The authors used again a cavity approach to estimate a critical point corresponding to $\Gamma_c \approx 1.78$ in our notation, but gave no appraisal of the error of this figure. Since these two models are conjectured to share the same thermodynamical properties, this estimate is expected to apply also to the spin glass model (3.1).

We consider the partition function of the system ($\beta = 1/T$)

$$Z = \operatorname{Tr} e^{-\beta H}, \tag{3.2}$$

assuming the dynamics is ergodic. This is almost certainly true if the system is coupled to a bath (which is the situation we consider here) but it is also probably true for an isolated system in a large region of parameters, including the paramagnetic region, although a complete analysis of this problem would require an analysis on the line of Refs [56, 57]. In particular we assume that taking the limit $T \to 0$ the region of parameters investigated is within the ergodic region (this is one of the reasons we do not push our analysis deep in the spin glass phase).

The transition at $\Gamma = 0$ is the classical transition (described in the previous chapter), as it is due solely to the Langevin forces generated by the interaction with the bath. This transitions is supposed to determine the characteristics of the whole line extending up to, but not including, $T = 0$. This can be understood using the Suzuki–Trotter approach in order to map the system defined by Eq. (3.1) to a classical systems with one additional dimension defined by an effective classical Hamiltonian H_{eff}. The details of this mapping are reviewed in Appendix A.2. By splitting the term βH into the sum of a number m of terms

$\Delta\beta H$ with $\Delta\beta \equiv \beta/m$ one can write

$$\begin{aligned} Z = \mathrm{Tr}\, e^{-\beta H} &= \sum_{s^{(0)}} \langle s^{(0)}| \Big(\prod_{i=1}^{m} e^{-\Delta\beta H} \Big) |s^{(0)}\rangle \\ &= \sum_{s^{(0)}} \cdots \sum_{s^{(m-1)}} \prod_{k=1}^{m} \langle s^{(k)}| e^{-\Delta\beta H} |s^{(k+1)}\rangle \\ &\approx \sum_{s^{(0)}} \cdots \sum_{s^{(m-1)}} e^{-\Delta\beta H_{\mathrm{eff}}(\vec{s})}. \end{aligned}$$

Here the set $\{|s^{(k)}\rangle\}_s$ is a product basis $|s^{(k)}\rangle = \bigotimes_i |s_i\rangle$ of the Hilbert space of the system where $|s_i\rangle = |\uparrow\rangle, |\downarrow\rangle$ and it is understood that $s^{(m)} = s^{(0)}$. The approximation is exact in the limit $m \to \infty$. The key point is the possibility of rewriting the bra-ket terms as

$$\langle s^{(k)}| e^{-\Delta\beta H} |s^{(k+1)}\rangle = e^{-\Delta\beta H^{(k)}(s^{(k)}, s^{(k+1)})},$$

where $H^{(k)}$ is a *classical* energy term defined on classical spins. Then $H_{\mathrm{eff}} \equiv \sum_k H^{(k)}$. In this way the system is given a transverse dimension of size $\beta = 1/T$, or more precisely $\sim J/T$. When one is close to the transition such that the critical system size exceeds this length the system can be "renormalized" back to the original RRG. This cannot be done when $T = 0$.

As a further point of interest, the ground state of the Hamiltonian of Eq. (3.1) is expected to be highly entangled somewhere along the $T = 0$ line. This is because the topological properties of the interaction structure of a local Hamiltonian such as those of Eq. (3.1) are known to affect the entanglement of its ground state. A very rough way of estimating the entanglement entropy of a region A, valid at least for the ground states of gapped Hamiltonians, is to assign a fixed contribution to each interaction that cross from A into the rest of the system. This implies that if many regions of the system are sparsely interacting with the rest of the system (*i.e.* if the spins inside of the region interact with only a few spins on the outside) then the entanglement will be comparatively low on average. On the other hand, if each spin participates in too many interactions then the entanglement in the ground state is suppressed as well, as the monogamy of entanglement [] prevents spins from being highly entangled with a large number of other spins. A Hamiltonian defined on an expander graph (such as a low-degree regular random graph) seems to lie halfway between these two extremal cases as these graphs are sparse but at the same time every subregion is highly-connected with the rest of the graph, and is therefore expected to be highly entangled.

Since decoherence is one of the main obstacles to building a fault-tolerant quantum computer, having a quantitative study of the theoretical amount of entanglement that is generated in an adiabatic path can provide a useful way of knowing the degree to which a quantum computer is performing a fully coherent adiabatic quantum algorithm. Consequently, one of the goal of this chapter is to study the entanglement dynamics of an *ideal* quantum annealing protocol, *i.e.* an actually adiabatic path where the transverse-field coupling constant Γ starts at a large value and is then quasi-statically decreased, in order to provide a benchmark for prospective in-depth studies of the entanglement dynamics of a real-life quantum annealer such as the D-Wave machine.

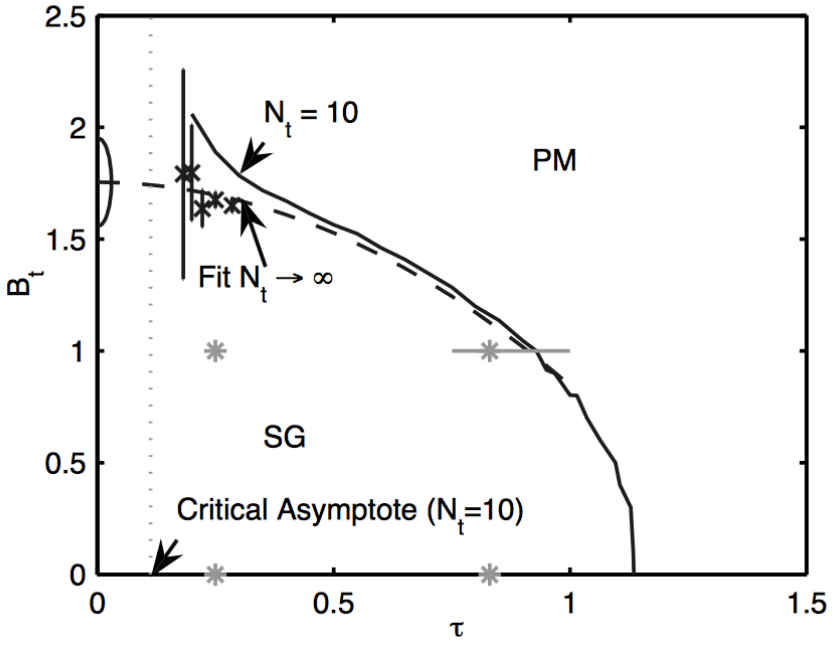


Figure 3.1: taken from Ref. [54]. Phase diagram of the model (3.1) obtained from a quantum version of the cavity method. This model exhibits a transition from a paramagnetic (PM) to a spin glass phase (SG) at small values of Γ and T (respectively equivalent to the B_t and τ coordinates in the figure). Note that the boundary between the two phases (solid black line) computed from the numerics is affected by a spurious deviation as the quantum cavity method become unreliable when one approaches the zero-temperature line.

3.2 Numerical Study

We numerically compute the Rényi entropy, quantum Fisher information, Edwards–Anderson order parameter and two-point correlation functions using exact diagonalization for small system sizes $N \approx 20$, and MC simulations for large systems sizes (up to $N = 140$). This allows us to obtain information about the thermodynamic-limit properties via finite-size scaling analysis.

3.2.1 Methods

Systems of small size ($N \leq 20$) are amenable to exact diagonalization (ED) methods, which means that the spectrum of the reduced density matrix is fully accessible.
We use the Lanczos algorithm to extract the ground state of H. This step constitutes the bottleneck of the whole procedure, as the Hamiltonian matrix is $2^N \times 2^N$ (albeit sparse), which effectively constrains the system size not to exceed $N \approx 20$ by too much. The reduced density matrix of half of the system, on the other hand, is only $2^{N/2} \times 2^{N/2}$, making it much easier to probe its full spectrum.

Another quantity we can probe via exact diagonalization is the quantum Fisher information of the ground state with respect to some observable $\hat{O}$. Following [36] we use the direction-averaged total spin operator, namely

$$\overline{F}_Q[\psi_0] \equiv \frac{F_Q[\psi_0, J_x] + F_Q[\psi_0, J_y] + F_Q[\psi_0, J_z]}{3},$$

where, for $a \in \{x, y, z\}$

$$J_a \equiv \frac{1}{2} \sum_{i=1}^{N} \sigma_i^a.$$

We noted in Section 1.4 (see Eq. (1.19)) that for a pure state $|\psi\rangle$ it holds that

$$F_Q[\psi, J_a] = 4\big(\langle\psi|J_a J_a|\psi\rangle - \langle\psi|J_a|\psi\rangle^2\big) = 4\,\mathrm{Var}(J_a)$$

so knowledge of the ground state $|\psi_0\rangle$ is sufficient for computing $\overline{F}_Q[\psi_0]$.
Where exact diagonalization stops being feasible (system sizes $N > 20$) one can compute thermal quantities numerically using the path-integral quantum Monte Carlo (PIMC) approach (see Appendix A.2). Expectation values of quantum observables $\mathcal{O}$ can then be computed in a straightforward manner by applying this quantum-to-classical map to the expression

$$\langle\mathcal{O}\rangle = \mathrm{Tr}(\mathcal{O}e^{-\beta H})/Z,$$

and then using standard Monte Carlo sampling on the resulting classical system. Ground-state properties are accessible in the limit $\beta \to \infty$. We use this standard PIMC approach when computing the quantum versions of the Edwards–Anderson order parameter of Eq. 2.7) and the spatial correlation functions of Eq. (2.8) (see Eqs. (3.6) and (3.7) later in this chapter).
However, one must take a somewhat different approach when computing entanglement entropies using Monte Carlo. Rényi entropies are *not* observables of the system (trivially, they are not linear operators) hence a naive PIMC is

inapplicable. In order to compute the Rényi entanglement entropy of a region A of the system we use the replica-approach described in Section 1.4.1 coupled with the algorithm of Ref. [58] (presented in Appendix A.4) that allows for computations of nonlinear quantities.

When computing the Rényi entropy using the Monte Carlo approach described before we set an inverse temperature of $\beta = 15$ in order to project the thermal state to the ground state and we use $m = 150$ timeslices for each replica so that the imaginary time discretization is given by $\Delta\beta = 1/10$. For all quantities we report in this thesis, we verified that the (small) systematic error due to the imaginary time discretization is within the statistical uncertainties. We also add a weak longitudinal field term $-h\sum_i \sigma_i^z$ with $h = 0.05$ to the Hamiltonian (3.1) since we noticed that the Monte Carlo convergence times are somewhat reduced if the $\mathbb{Z}_2$ symmetry is partially lifted. It is important to stress that this small longitudinal field does not qualtitatively affect the system properties, in particular its glassy nature at small Γ values. In fact, this glassy nature makes the PIMC computations of the Rényi entropy rather computationally expensive, in particular for small Γ values and large system sizes, where the Edwards-Anderson order parameter is large. The (small) effect that this field has on the Rényi entropy values is expained in Subsection 3.2.3. In order to assign an error bar to the result for each disorder realization we performed ten independent simulations starting from a different initial configuration (the same for all replicas). In this case, the value of $S_A^{(2)}$ for each realization of disorder is obtained by averaging over the ten results. For the Edwards–Anderson order parameter and the correlation functions we use $\beta = 40$ and $\Delta\beta = 1/8$.

We emphasize that our analysis focuses on the regime of large and intermediate values of $\Gamma \gtrsim \sqrt{K} > 1$ where K is the connectivity of the RRG. Here, for the finite system size we consider, the problems due to the slowdown of the Monte Carlo dynamics in the glassy phase are not overwhelming. More details on all these techniques are contained in the Appendix A.2.

3.2.2 Average over the disorder

Our system contains disorder and our numerical methods only work on a fixed realization of disorder at a time. The natural way of taking a disorder average is then to generate a large number of realizations, compute the desidered quantity for each realization and then take the average of the results. When computing these average values for a fixed system size N we have to consider two distinct disorder variables: *i)* the topological structure of the interaction graph $G = (V, E)$ and *ii)* the couplings J_{ij} in the interaction Hamiltonian. A specific instance of the simulation is defined by the values assigned to these two variables. To create such an instance $i_N = i_N(G, J_{ij})$ for a system of N spins we generate an N-vertex regular random graph using the Steger and Wormald algorithm [59] and then we randomly assign a coupling $J_{ij} = \pm 1$ to each edge of the graph, with equal probability.

In the case of the Rényi entropy an instance is defined by *three* variables $i_N = i_N(G, J_{ij}, A)$, since one has to take into account also the region A of which we compute the Rényi entropy. This region is generated according to the Random Region protocol described in Fig. (3.2).

RANDOM REGION

We select a connected region A of $N/2$ vertices in the following way

1) Select a vertex v uniformly at random and define $A \equiv \{v\}$.

2) Given the region A, select uniformly at random a vertex u from the boundary ∂A of the region A. Then update the region $A \mapsto A' \equiv A \cup \{u\}$.

3) Repeat point 2) until the desired size of A is reached.

Figure 3.2: Protocol for the generation of a random region A.

3.2.3 Results

Rényi entropy — Our first goal is understanding the thermodynamic-limit behaviour of the disorder-averaged Rényi entropy of order $\alpha = 2$ ($\overline{S}_A^{(2)}$, in the following abbreviated as Rényi-2) on the $T = 0$ line of the (T, Γ)-phase diagram of the Hamiltonian (3.1). The choice $\alpha = 2$ was made because it is the least demanding in terms of computational resources. First we compute the finite-size values $\overline{S}_A^{(2)}(N)$ and then we study the limit $N \to \infty$.
Small-sizes results obtained by exact diagonalization suggest a volume-law scaling for all values of Γ (Figs. 3.3,3.4), which is confirmed by the Monte Carlo results (Figs. 3.5,3.7). Comparing the QMC and the ED results we note that the weak longitudinal field we use in the Monte Carlo simulations reduces the numerical value of the Rényi entropy in the critical region while leaving the positions of the peaks essentially unaffected. Moreover, we verified that ED and QMC results agree on system sizes where both can be applied when the weak longitudinal field is included also in the ED calculations. We see that, for each system size N, the $S_A^{(2)}$ curve attains a peak. These peaks of the entanglement entropy have been associated to criticality in quantum phase transitions [60], where the position of the maximum in the thermodynamic limit is the critical point of the transition. In order to extract this value from our finite-size data we define, for each finite N, the estimator $\Gamma_c(N)$ as the point where the $S_A^{(2)}(N)$ curve attains its maximum and then we study the limit as $N \to \infty$.
We approximate the peaks by fitting the numerical data with parabolic curves and then we take the vertices of the parabolas to be the finite-size estimators $\Gamma_c(N)$. Then we use a simple $1/N$ fitting ansatz to extract $\Gamma_c \equiv \lim_{N\to\infty} \Gamma_c(N)$ by looking at the convergence of these peaks in the thermodynamic limit

$$\Gamma_c(N) \equiv \Gamma_c + \frac{\Delta\Gamma_c}{N}. \tag{3.3}$$

We obtain a critical point of $\Gamma_c = 1.84 \pm 0.02$ (Fig. 3.6). By considering both the critical scaling of $S_A^{(2)}$ and the finite-size shift on $\Gamma_c(N)$ we propose the finite-size scaling (FSS) ansatz

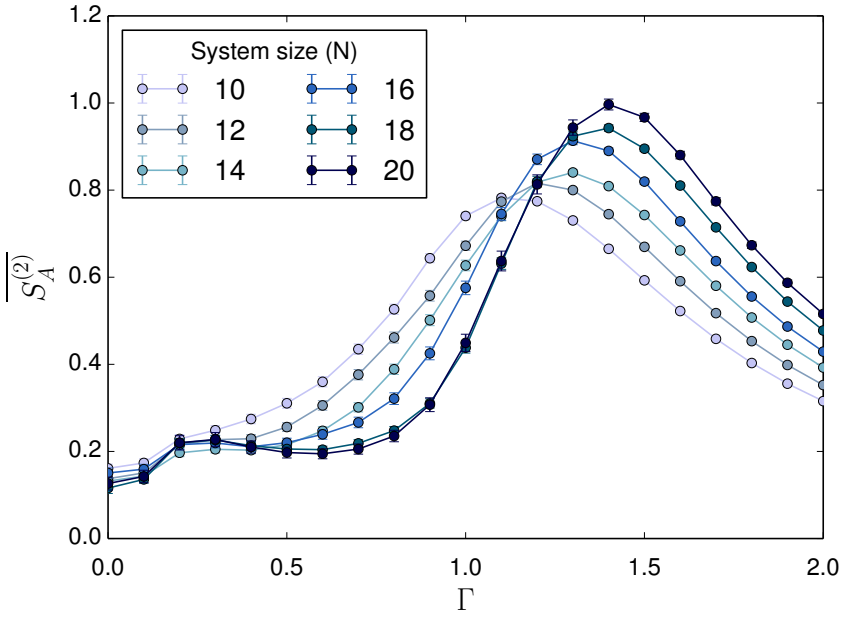


Figure 3.3: Plot of the average value of the Rényi entropy $\overline{S_A^{(2)}}$ as a function of the transverse field strength Γ, for different sizes of the system (exact diagonalization results).

$$S^{(2)}(N,\Gamma) = \tilde{s}\Big(\Gamma - \Gamma_c(N)\Big)(aN+b), \tag{3.4}$$

which gives the data collapse shown in Fig. (3.7). As explained above, the convergence times of the PIMC calculations increase when going deep into the glassy phase, where the Edward-Anderson order parameter is large; this sets some limitations to the regime of Γ and N values that we can explore.

Quantum Fisher Information — We compute the Quantum Fisher Information (QFI) by exact diagonalization for system sizes $N \leq 20$. The QFI curves have a shape reminiscent of the entanglement entropy curves, including the peak at $\Gamma_c(N)$. A FSS analysis of the peak leads to Fig. 3.9, confirming a linear size scaling with linear coefficient $a = 0.85(1)$.

It can be shown that the condition (basically a refinement of Eq. (1.20))

$$\overline{F}_Q[\psi] > \frac{1}{3}\left[s(k^2+2k-\delta_{k,1}) + r^2 + 2r - \delta_{r,1}\right], \tag{3.5}$$

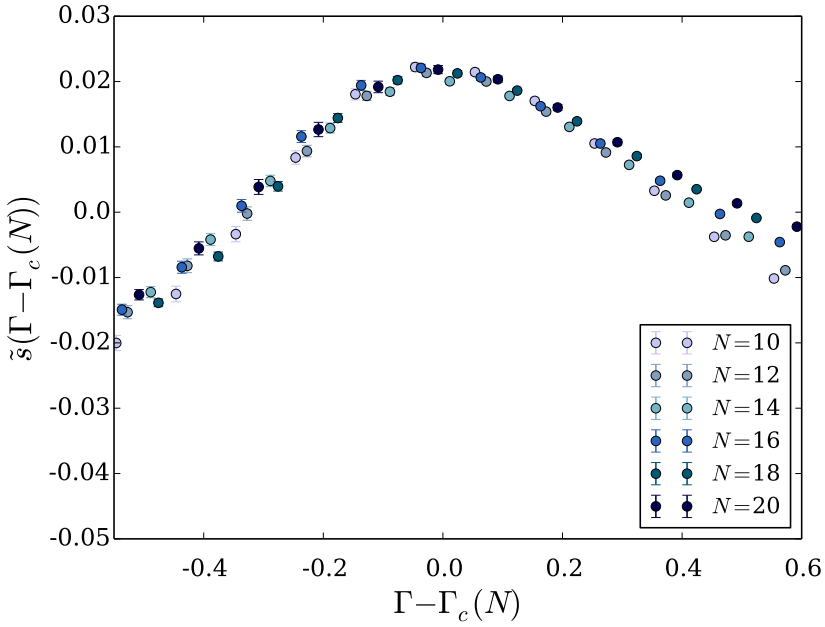


Figure 3.4: Data collapse for the finite-size scaling of the data of Fig. 3.3.

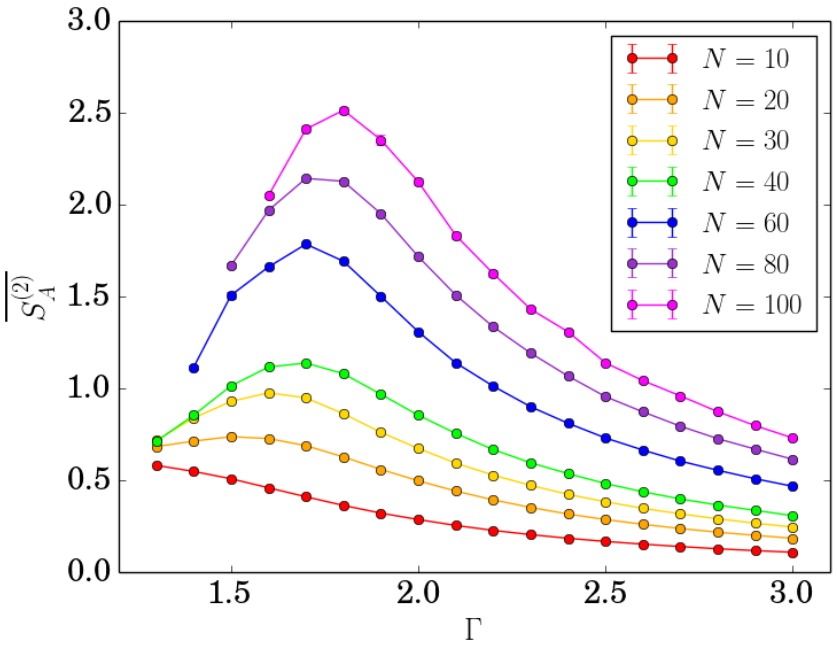


Figure 3.5: Plot of the average value of the Rényi entropy $\overline{S_A^{(2)}}$ as a function of the transverse field strength Γ, for different sizes of the system (quantum Monte Carlo results).

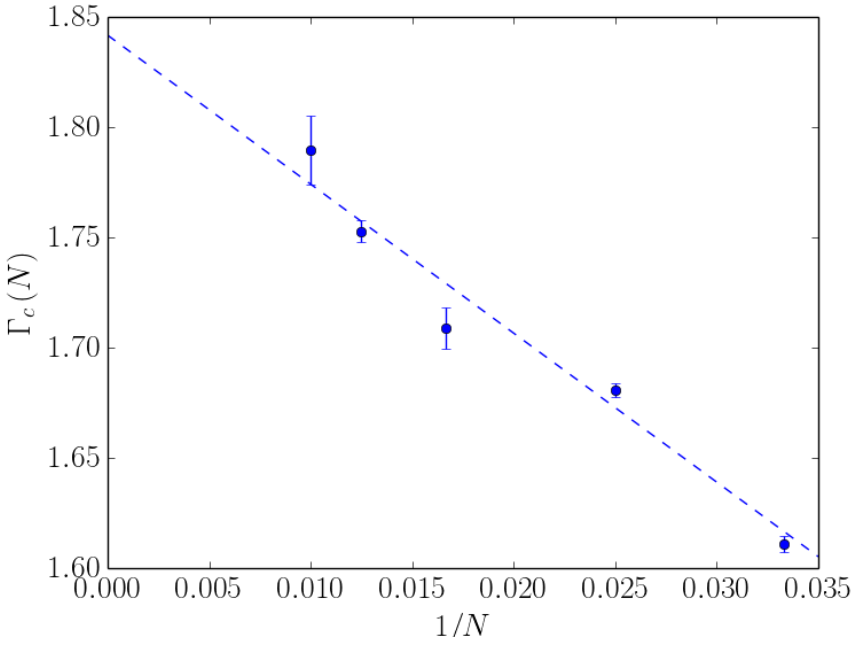


Figure 3.6: Estimated maxima of $S_A^{(2)}$ are fitted with a $1/N$ scaling behaviour.

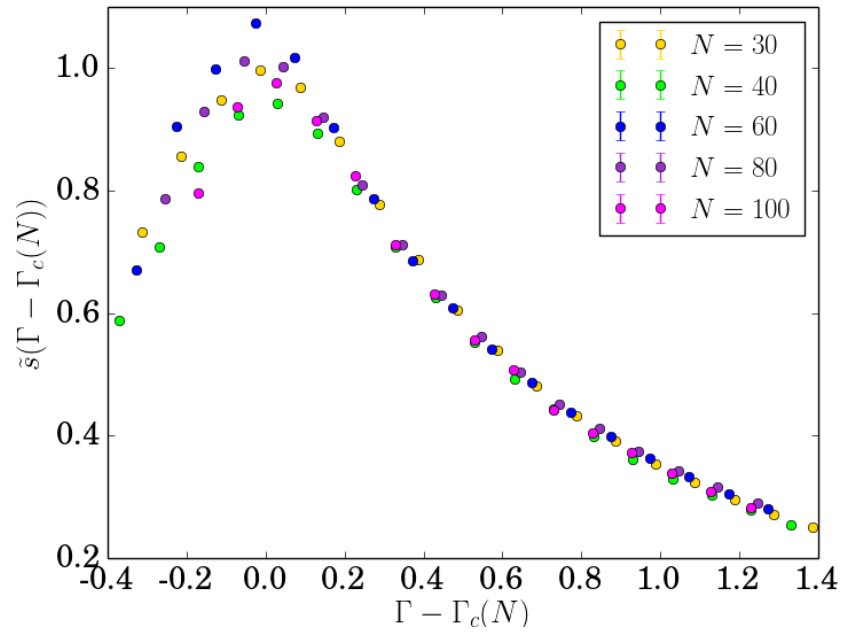


Figure 3.7: Data collapse for the finite-size scaling of the Monte Carlo data. Finite-size estimates to the critical point Γ_c are given by the fit $\Gamma_c(N) = \Gamma_c + \Delta\Gamma_c/N$.

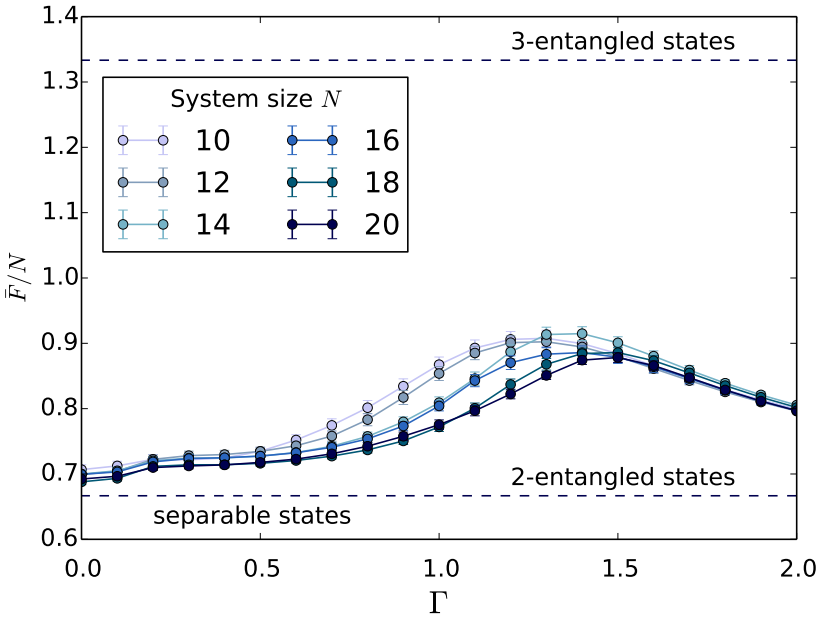


Figure 3.8: Disorder-averaged ground-state quantum Fisher information density $\overline{F}_Q/N$ as a function of the transverse field strength Γ. The lower dashed line represents the separability criterion violation: states above it are entangled. States above the upper dashed line are at least 3-entangled.

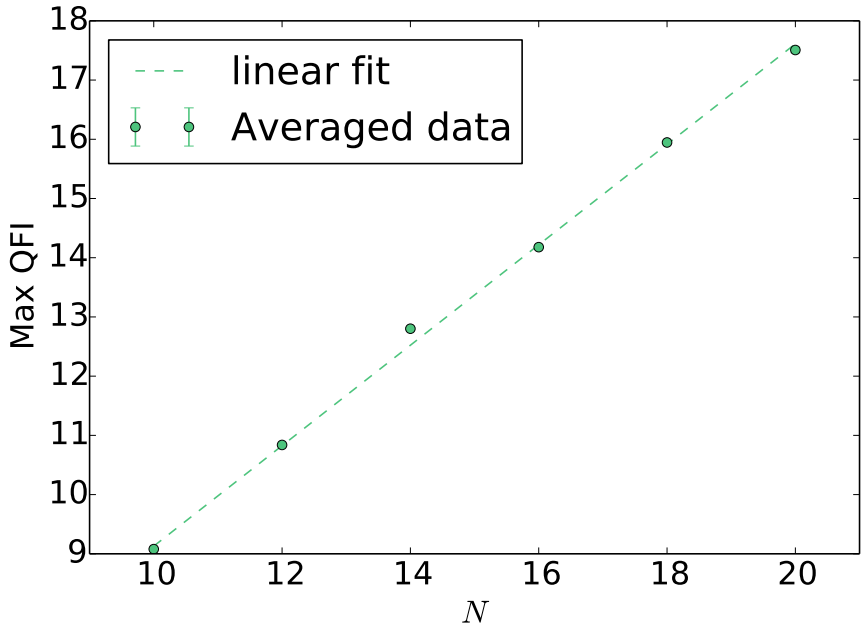


Figure 3.9: Linear fit of the ground state QFI peak.

where $s = \lfloor N/k \rfloor$ and $r = N - sk$, implies that the state $|\psi\rangle$ contains multipartite entanglement between at least $k+1$ spins [36, 37], *i.e.* it cannot be written as $|\psi\rangle = \bigotimes_i |\psi_i\rangle$ where each $|\psi_i\rangle$ is a state of $n_i \leq k$ spins.
Note that for N even, proving 3-entanglement from Eq. (3.5) requires $\overline{F}_Q[\psi_0]/N \geq 4/3$, which FSS suggests will never be satisfied (since $0.85 < 4/3$, see Fig. 3.8 and 3.9): the entanglement in the ground state seems to be limited to two-spins states. We emphasize that the QFI $F_Q[\cdot, J]$ is dependent on the specific choice of the generator J so, strictly speaking, the previous results are to be taken as *lower bounds* to the multipartite entanglement present in the ground state. The actual value is obtained by maximizing $F_Q[\cdot, J]$ over all possible choices of J, as explained in Section 1.4. We are unable to solve this maximization problem but our choice of generator seems to be a natural one when considering only one-spin operators.

Edwards–Anderson order parameter — In order to characterize the glassy properties on the low Γ phase, and also to cross-check our estimate for the critical point Γ_c, we compute the quantity

$$q_{EA} \equiv \frac{1}{N} \sum_{i=1}^{N} \langle \sigma_i^z \rangle^2, \tag{3.6}$$

in the ground state of the Hamiltonian of Eq. (3.1))over a 3-regular random graph. While the $\mathbb{Z}_2$ symmetry of Eq. (3.1) implies $q_{EA} = 0$ for all finite sizes, by running a Monte Carlo simulation for a finite time one can effectively reproduce the spontaneous breaking of the symmetry that is encountered in the

thermodynamic limit, as the probability of having $\mathbb{Z}_2$-transitions vanish in the limit $N \to \infty$. For each specific realization of the disorder $i_N = i_N(G, J_{ij})$ we obtain the quantity of Eq. (3.6) by computing each value of $\langle \sigma_i^z \rangle$ via Monte Carlo simulations. Then we take the average over the disorder obtaining a curve $q_{EA}^{(N)}(\Gamma) \equiv q_{EA}(\Gamma, N)$ at fixed size N (results are shown in Fig. 3.10). The critical point $\Gamma = \Gamma_c$ is then the point of singularity of the curve

$$q_{EA}(\Gamma) = \lim_{N\to\infty} q_{EA}^{(N)}(\Gamma).$$

We observe strong finite-size effects and most curves have smooth transitions from a region where they are zero (within statistical error) to a region where they attain positive values. Therefore, for any fixed size N, our finite-size estimates of the critical point $\Gamma_c(N)$ are obtained in the following way. First we shift all of the curves horizontally so that they fall on top of each other, as in Fig. (3.11). Then we take a linear fit of the accumulated data around the point where q_{EA} starts being finite and obtain a slope s. Finally we take linear approximations to each of the $q_{EA}^{(N)}$ curves in Fig. (3.10) using the fixed slope s. The values $\Gamma_c(N)$ are defined as the x-intercept of these new linear fits. The ansatz

$$\Gamma_c(N) = \Gamma_c + \frac{\Delta\Gamma_c}{N}$$

for these points gives $\Gamma_c = 1.82 \pm 0.02$. We get an excellent agreement with the critical point we extracted from the study of the Rényi entropy. We see this as an indication that the QPT we detected using the Rényi entropy is exactly the glassy phase transition of the model.

Connected correlations — We compute the connected correlation function

$$C_{ij} \equiv \langle \sigma_i^z \sigma_j^z \rangle - \langle \sigma_i^z \rangle \langle \sigma_j^z \rangle \tag{3.7}$$

of the Hamiltonian (3.1) as follows: for each realization of disorder we randomly choose a central spin i on the interaction graph and we compute all connected correlation functions C_{ij} for all sites j in the system. Then for any integer r we define the mean correlation at distance r, $C_{\text{mean}}(r)$, and the maximal correlation at distance r, $C_{\max}(r)$, by taking respectively the average and the maximum of (the absolute value of) the correlations among all sites j that are r steps away from the central spin i in the distance of the graph $G = (V, E)$

$$\begin{aligned} C_{\text{mean}}(r) &\equiv \text{mean}\{\, |C_{ij}| : j \in V, \text{dist}(i,j) = r\}, \\ C_{\max}(r) &\equiv \max\{\, |C_{ij}| : j \in V, \text{dist}(i,j) = r\}, \end{aligned}$$

then we average these quantities over many realizations of disorder to get the average mean correlation $\overline{C}_{\text{mean}}(r)$ and the average maximal correlation $\overline{C}_{\max}(r)$ as functions of the distance r. We find that both $\overline{C}_{\text{mean}}$ and $\overline{C}_{\max}$ follow a stretched-exponential behaviour

$$\overline{C}_{\text{mean,max}} \sim e^{-(r/\xi)^a}$$

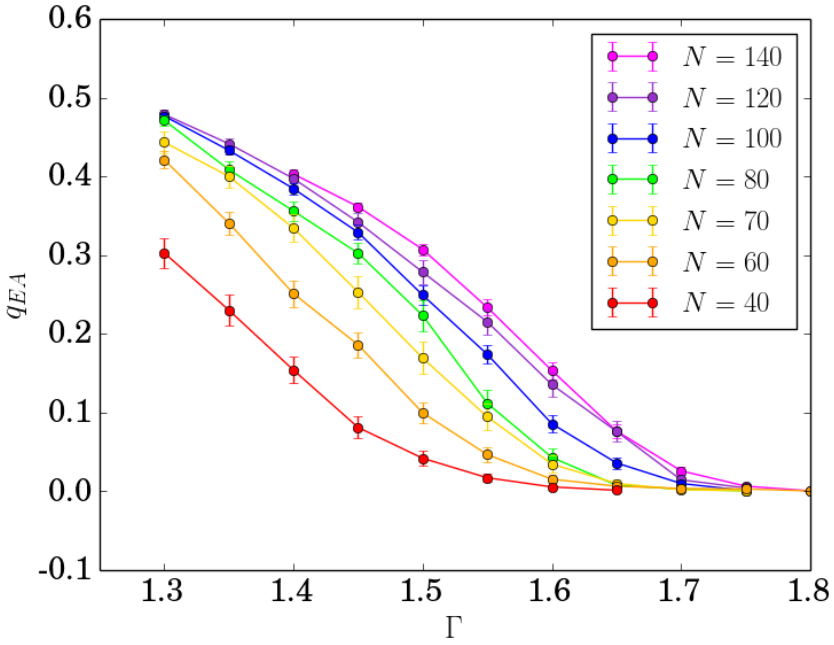


Figure 3.10: Disorder-averaged Edwards–Anderson order parameter q_{EA} as a function of the transverse field strength Γ, for different system sizes. The finite-size effects induce a smoothing-out of the curves that disappears as $N \to \infty$.

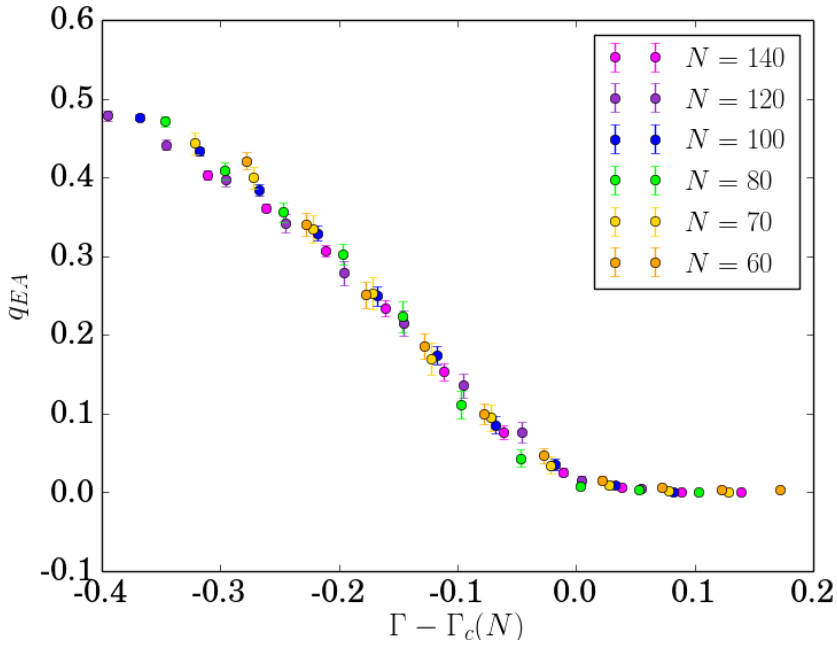


Figure 3.11: Data collapse of the curves of Fig. (3.10). The scaling for $\Gamma_c(N)$ is shown in Fig. 3.15.

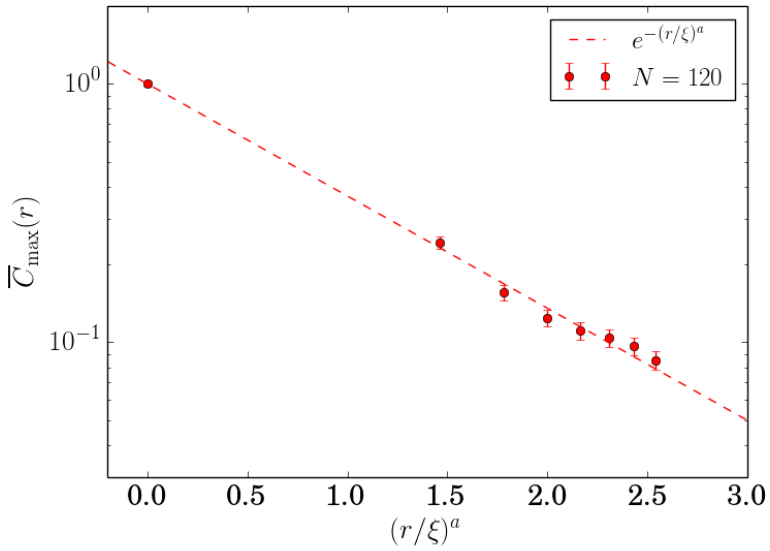


Figure 3.12: Stretched expenential fit for the disorder-averaged maximal correlation $\overline{C}_{\max}$ for system size $N = 120$ and $\Gamma = 1.50$. Fitting parameters are $a = 0.28$, $\xi = 0.25$. Other values of N and Γ give qualitatively similar results, with different values for the fitting parameters a, ξ.

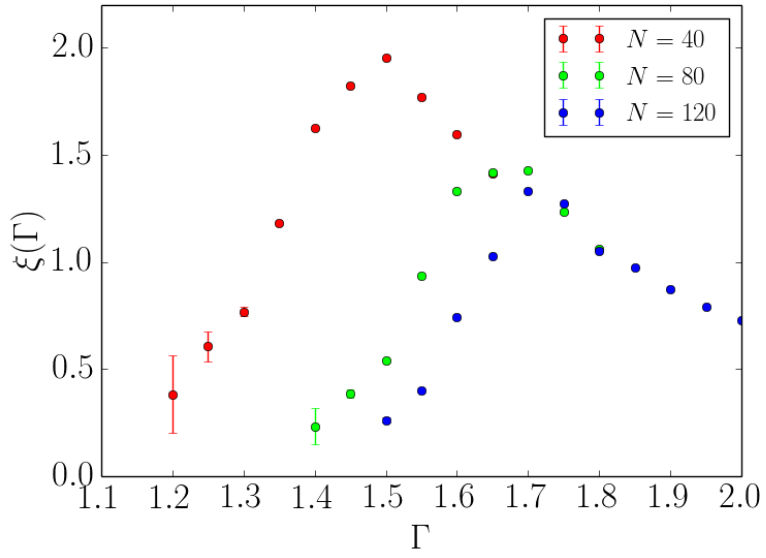


Figure 3.13: Correlation lengths $\xi(\Gamma)$ obtained from the stretched-exponential fit to the numerical results of $\overline{C}_{\max}$. Finite-size estimators $\Gamma_c(N)$ for the critical point are defined as the vertices of parabolic fits to the curves.

at any transverse field value Γ (see Figure 3.12). Stretched exponentials are usually encountered in disordered systems when taking the disorder average of an exponentially-decaying quantity Q that has different exponential decay constants ξ_i for each realization of disorder i, *i.e.* $Q_i(r) \sim e^{-r/\xi_i}$. Therefore we argue that our model shows a distribution of correlation lengths ξ_i between different disorder realizations. By defining $\Gamma_c(N)$ as the value of Γ where the curve $\xi(\Gamma)$ for size N attains its maximum, we note that for both quantities $\overline{C}_{\text{mean}}$ and $\overline{C}_{\max}$, the critical correlation lengths $\xi_{\Gamma_c}(N)$ we obtained from the fits are *decreasing* with the system size N, and converge to a finite value in the thermodynamic limit (Fig. 3.14).

Critical Point — Summarizing our results, we have computed four independent ways of estimating the critical point of the glassy phase transition, *i.e.* four sets of finite-size estimators $\Gamma_c(N)$ that are free of the systematic error that affected previous cavity method approaches. Each set was extracted from the study of a different physical quantity whose behaviour is known to be affected by criticality. Finite-size corrections disappear as $1/N$ and in the thermodynamical limit all estimates seem to converge to a critical point $1.82 \leq \Gamma_c \leq 1.85$, as shown in Fig. (3.15). This estimate is compatible with the previous ones in Refs. [54, 55].

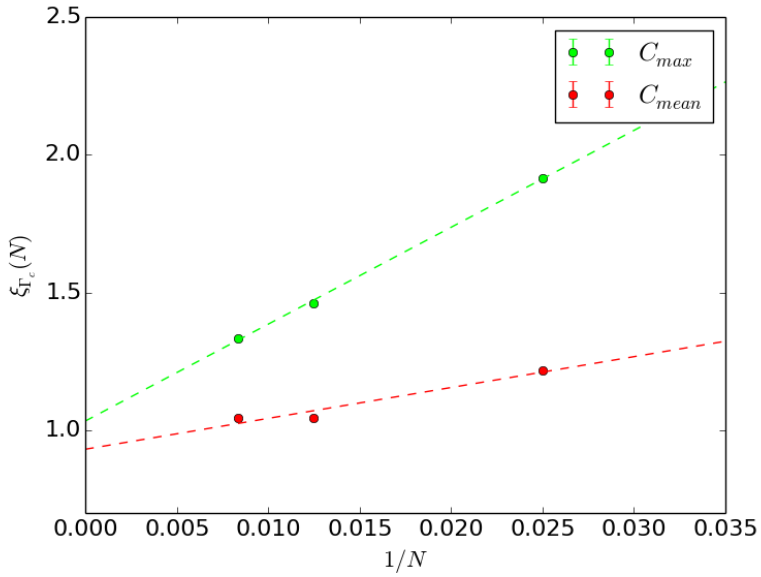


Figure 3.14: $1/N$ fit of the critical correlation lengths obtained from $C_{\max}$ and C_{mean}.

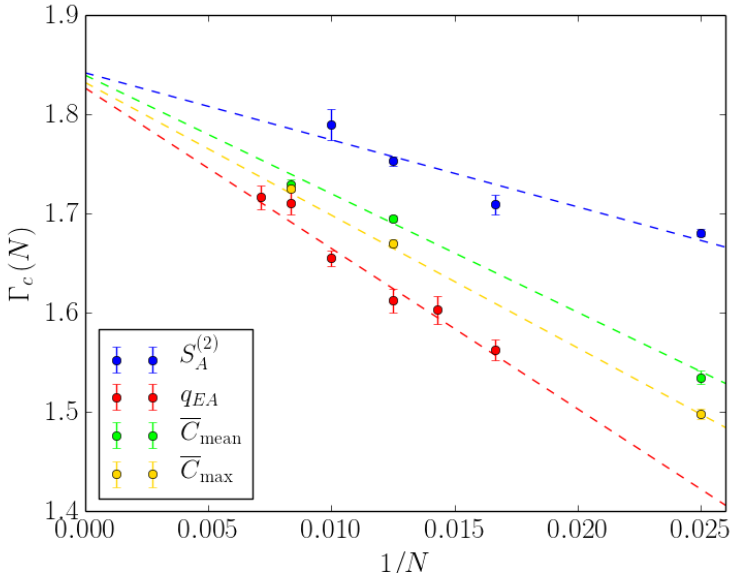


Figure 3.15: $1/N$ fits of the finite-size estimators $\Gamma_c(N)$ for the critical point obtained from all the quantities considered in this work. While finite-size corrections differ between quantities, all seem to indicate a critical point $1.82 \leq \Gamma_c \leq 1.85$.

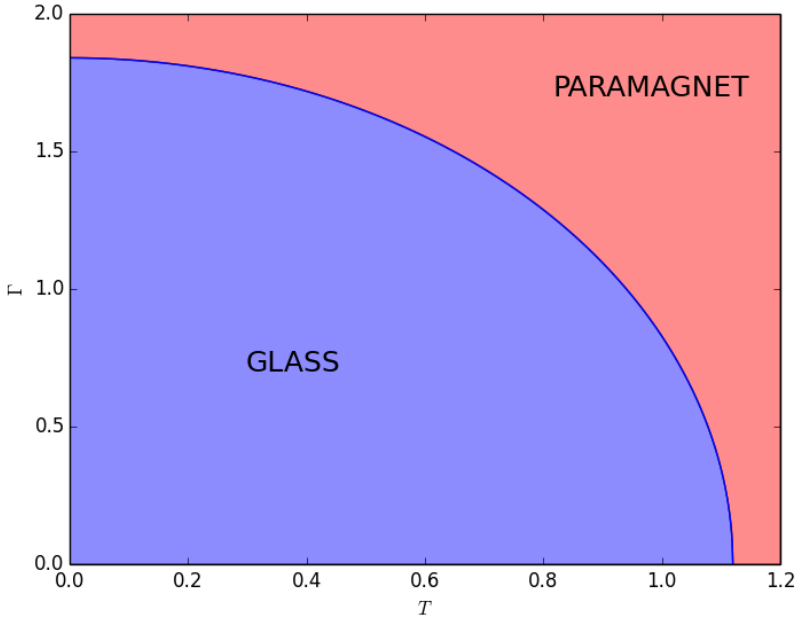


Figure 3.16: Phase diagram of the model (3.1) as computed from our numerical results.

3.3 A mean-field model of the transition

The fact that $\Gamma_c/J \gtrsim 1$ (and, as we will see, that it grows with K), should suggest us that a better starting point for analyzing the transition is $\Gamma = \infty$ rather than $\Gamma = 0$, which has been the method adopted in previous investigations of the glassy phase.

We are going to show that a simple mean field theory, obtained by dressing the excitations around $\Gamma = \infty$, gives a transition point quite close to the unbiased prediction obtained from the Monte Carlo simulations and predicts two main things: a) the scaling of the critical point $\Gamma_c \propto \sqrt{K}$ and b) the exponential decay of correlations inside the paramagnetic phase. Both facts can be deduced from other thermodynamic considerations (the first one for example by comparing the order of the two terms in the Hamiltonian for large K), but our derivation will highlight some properties of the dynamics of the excitations as seen as quasiparticles. In particular, by connecting the $\sqrt{K}$ factor to *delocalization* of the excitations we exclude the possibility that the glass transition coincides with the MBL transition. We will return on this point after Eq. (3.20).

If we take $J \equiv |J_{ij}|$ to be the perturbative parameters, then the "free" Hamiltonian is the transverse-field term $H_0 \equiv -\Gamma \sum_i \sigma_i^x$. By shifting the ground state energy $E_0 = -\Gamma N$ to zero, its spectrum is given by

$$\mathrm{Sp}(H_0) = \{0, 2\Gamma, 4\Gamma, \ldots, 2N\Gamma\}.$$

Its unique ground state $|0\rangle$ (where we denote $\sigma^x|\rightarrow\rangle = |\rightarrow\rangle$)

$$|0\rangle \equiv \bigotimes_{i=1}^{N} |\rightarrow\rangle_i,$$

is taken as a pseudovacuum of quasiparticles. The operators σ_i^z create an excitation on top of the ground state

$$|i\rangle \equiv \sigma_i^z|0\rangle = |\leftarrow\rangle_i \otimes \Big(\bigotimes_{j\neq i} |\rightarrow\rangle_j\Big),$$

which is interpreted as a state containing a quasiparticle at site i. These form the N-fold degenerate eigenspace of H_0 with energy 2Γ. Additional applications of the $\{\sigma_j^z\}$ operators either move the state upwards in the spectrum, creating states with two, three, or more quasiparticles $|i,j\rangle, |i,j,k\rangle, \ldots$, or downwards, by annihilating existing quasiparticles.
The perturbed Hamiltonian is

$$H(J) = H_0 + JV$$

where V is the dimensionless spin-glass term $V = -(1/J)\sum_{\langle i,j\rangle} J_{ij}\sigma_i^z\sigma_j^z$
First-order perturbation theory gives a null correction to the ground-state energy

$$E_0(J) = JE_0^{(1)} = 0$$

since $JE_0^{(1)} = \langle 0|JV|0\rangle = 0$. Instead, the degenerate band of one-particle states is split by the perturbation into distinct levels

$$E_m(J) = 2\Gamma + JE_m^{(1)} \tag{3.8}$$

for $m = 1, \ldots, N$, where JE_m are eigenvalues and $\{|\phi_m\rangle\}$ are the eigenvectors of the operator $\tilde{V}$, which is the perturbation operator V restricted to the (unperturbed) one-particle subspace. The direct computation of the $E_m^{(1)}$ requires the diagonalization of an N by N matrix which, we will show, is a hopping matrix on the RRG.
Let us start by noticing that the term $\sigma_i^z\sigma_j^z$ applied to a quasiparticle state $|i', j', k', \ldots\rangle$ can affect it in one of three ways: (i) it can create a pair of adjacent quasiparticles provided the sites i and j are devoid of them, (ii) it can move a quasiparticle from site i to site j provided site j is empty and site i is occupied (or vice versa, swapping the role of i and j), or (iii) it can annihilate a pair of adjacent quasiparticles sitting on sites i and j.
Note also that in the one-particle subspace annihilation processes cannot happen, while creation processes map a state into the three-particle subspace, which is orthogonal to the one-particle subspace. Therefore, only the hopping processes give a contribution to Eq. (3.8). For any state $|\psi\rangle$ in the one-particle subspace, the action of $J\tilde{V}$ is then equivalent to that of H_{hop}, the Hamiltonian of a particle hopping on the same graph G, with disordered hopping constants $J_{ij} = \pm J$

$$H_{\text{hop}} = -\sum_{\langle i,j\rangle} J_{ij}\Big(|i\rangle\langle j| + |j\rangle\langle i|\Big).$$

For contrast, consider the Hamiltonian $H_{\rm hop}^{(\rm hom)} = -J\sum_{\langle i,j\rangle}\Big(|i\rangle\langle j| + |j\rangle\langle i|\Big)$ of a particle hopping on the graph G with homogeneous hopping coefficients $J > 0$. Even though solving for the spectrum of the latter Hamiltonian is difficult for any finite graph G, in the thermodynamic limit, our RRG G is the Bethe lattice and so spectral properties of this model can be computed exactly using an iterative method (see *e.g.* [61, 62] and Fig. 3.17). In particular, its spectral density is known to be supported on the set $[-2J\sqrt{K}, 2J\sqrt{K}]$, where K is the constant connectivity of the graph. We give the proof here: one starts by writing down iteration equations for the diagonal Green's function $G_i \equiv \langle i|(E - H_{\rm hop})^{-1}|i\rangle$ at the site i for generic complex E

$$G_i = \left(E - \sum_{j\in\partial i} J_{ij}^2 G_j^{(c)}\right)^{-1}, \tag{3.9}$$

where the cavity Green's function $G_j^{(c)} = \langle j|(E - H_{\rm hop}^{(i)})^{-1}|j\rangle$ is the Green's function of the operator $H_{\rm hop}^{(i)}$ obtained from the Hamiltonian $H_{\rm hop}$ by removing the hopping terms associated to the edges incident to the site i. On the Bethe lattice the removal of these terms splits the system into $K+1$ isomorphic disconnected components that can be considered independently. Each component is an infinite rooted tree with a branching factor of K. These trees are isomorphic to each of their infinite descending subtrees rooted at any of their vertices. By writing the iteration equation (3.9) for $G_j^{(c)}$ one gets

$$G_j^{(c)} = \left(E - \sum_{k\in\partial j} J_{jk}^2 G_k^{(c')}\right)^{-1}, \tag{3.10}$$

where $G_k^{(c')}$ is a second-step cavity Green's functions obtained from the $G_j^{(c)}$ Hamiltonian by further removing the hopping terms associated to the edges incident to j. By solving (3.10) and plugging the result in (3.9) one recovers G_i and from G_i the spectral density

$$\rho_i(E) = \frac{1}{\pi}\lim_{\mathrm{Im}\,E\to 0^+} G_i(E).$$

One can find $P(\mathrm{Im}\,\Sigma_i)$ by $\Sigma_i(E) = E - G_i^{-1}$ and taking the limit $\mathrm{Im}\,E \to 0$. As we know from [63], the distribution of $\mathrm{Im}\,\Sigma$ will tend to a delta function for *delocalized states* and to a long-tailed distribution for *localized states*. However we can sidestep all this procedure recognizing that in all the equations only J_{ij}^2 appears and therefore the case $J_{ij} = \pm J$ is *identical* to the case $J_{ij} = J$ constant, which has only delocalized states. The latter statement follows from the observation that, given the constant connectivity of the graph and the constant value of the hopping J the distribution of $G^{(c)}$ must be a delta function centered on the solution of the deterministic equation (3.9)

$$G^{(c)} = \left(E - KJ^2 G^{(c)}\right)^{-1}, \tag{3.11}$$

which gives

$$G^{(c)} = \frac{2}{E - \sqrt{E^2 - 4KJ^2}}. \tag{3.12}$$

By inserting (3.12) into (3.9) one gets

$$\begin{aligned} G_i^{-1} &= E - (K+1)J^2 G^{(c)} \\ &= E - \frac{2(K+1)J^2}{E - \sqrt{E^2 - 4KJ^2}} \end{aligned}$$

so

$$\Sigma_i = \frac{2(K+1)J^2}{E - \sqrt{E^2 - 4KJ^2}}$$

irrespective of i and therefore all the eigenstates are *delocalized*[1].

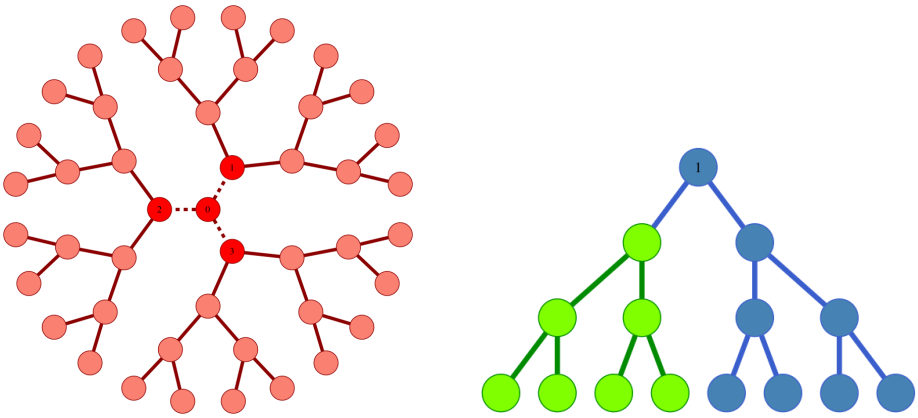


Figure 3.17: (Left) The diagonal Green's function G_i at the point $i = 0$ on a 3-regular Bethe lattice can be computed once the values $G_j^{(c)}$ are known at the points $j = 1, 2, 3$ (shown in red). $G_j^{(c)}$ is the Green's function, computed at the site j, of the operator obtained from the Hamiltonian H_{hop} by removing the hopping terms terms associated to the dashed edges of the graph. (Right) Each disconnected component thus obtained is an infinite binary tree that is isomorphic to all its infinite descending binary subtrees (one such subtree is shown in green).

We take the greatest lower bound of the support of the spectral density $\rho(E)$ as the energy $E^{(1)} = -J\sqrt{K}$ of the ground state in the thermodynamic limit. If we neglect processes of $O(J^2/\Gamma^2)$ which dress both the ground state and the excited states we get a crossing when $E^{(1)} + 2\Gamma = 0$, or

$$\Gamma_c^{(1)} = J\sqrt{K} \simeq 1.41J. \tag{3.13}$$

This is an underestimate, but quite a good one, of the critical point as observed in the numerics $1.82 \leq \Gamma_c \leq 1.85$. Notice also that this scaling of Γ_c with K is important as the limit $K \to \infty$ should reproduce the fully connected, Sherrington-Kirkpatrick model [64, 65, 66], by scaling $J \to J/\sqrt{K}$. This scaling keeps the transition temperature T finite and we find here that the independent particle picture also returns a finite Γ.

[1] we will more closely examine localization in the next chapter.

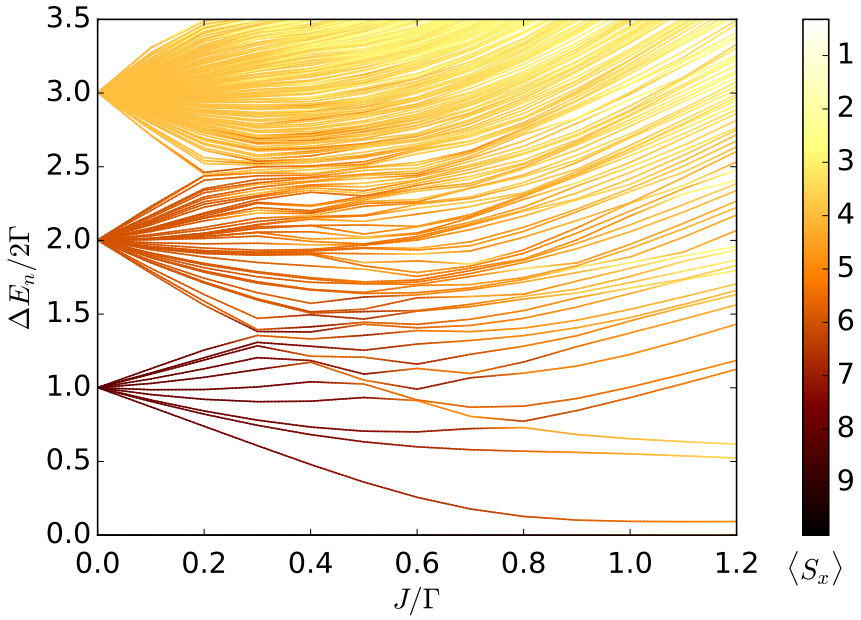


Figure 3.18: Perturbative lifting of the degeneracy of the bands in the spectrum of the transverse-field Hamiltonian H_0 (at fixed Γ) as J is increased from zero. Color code denotes the average value of $S_x \equiv \sum_i \sigma_i^x$ of the state associated with the energy level.

We now try to go beyond the independent particle approximation. By going to higher n sectors we need to solve the problem with n particles which, to first order in J/Γ, are hard-core bosons. This is a complicated problem (which, by the way, is equivalent to replacing $\sigma_i^z\sigma_j^z \to \sigma_i^+\sigma_j^-$). If $n \ll N$, as a first approximation we can assume the particles as non-interacting.

First, note that as soon as n is finite when $N \to \infty$ we can conclude that all such few-particle states are delocalized. As n/N becomes appreciable so does the interaction between the particles, we shall conjecture that the introduction of a small interaction between the particles (which is of $O(J^2/\Gamma)$), irrespective of its attractive or repulsive nature, does not *localize* excitations (this is true even in the presence of bound states) and the phase remains delocalized. The value of Γ/J where this breaks down we cannot predict without considering quantitatively the interaction between the particles.

As for the critical point of the transition, note that the non-interaction assumption gives for the ground state of the n-th sector $E_0^{(n)} = -2nJ\sqrt{K}$, and by matching $E^{(n)} + 2n\Gamma = 0$ we find the very same result of Eq. (3.13).
In the spirit of mean-field theory we can add the interaction of the particles by observing that, when considering only two-particles processes, the process $J\sigma_i^z\sigma_j^z|\leftarrow,\leftarrow\rangle = |\rightarrow,\rightarrow\rangle$ which reduces by 4Γ the energy of the state, can be interpreted as a zero-range attraction potential energy $V = -4\Gamma$ when two particles are on top of each other. This correction to the energy of the ground state goes in the right direction, increasing the value of Γ_c. In fact, we have an expected energy per particle $V\rho/2 = -2\Gamma\rho$ where $\rho = n/N \ll 1$ is the probability to find two particles on the same site, assuming the particles are indeed delocalized

$$E^{(n)}/n = -2J\sqrt{K} - 2\Gamma\rho, \tag{3.14}$$

and therefore the equation defining the critical point is $-2J\sqrt{K} - 2\Gamma\rho + 2\Gamma = 0$ so

$$\Gamma_c^{mf} = J\sqrt{K}(1 + \rho - O(\rho^2)). \tag{3.15}$$

The modification is in the right direction: increasing the density $\rho = n/N$, the value of Γ_c at the critical point increases; moreover the $\sqrt{K}$ scaling is preserved. Higher and higher n sectors cross the ground state at larger Γ. Of course this approximation is reliable only if $\rho \ll 1$ (agreement with the observed $1.82 \leq \Gamma_c \leq 1.85$ requires $\rho \simeq 0.3$, close to the center of the band $\rho = 1/2$, although one cannot claim $K = 2$ to be in the correct large-K regime).

Let us now see how to connect the computed spin–spin correlation function with the propagator of the particle on the Bethe lattice. From the definition

$$C(i,j) = \langle\Psi_0|\sigma_i^z\sigma_j^z|\Psi_0\rangle - \langle\Psi_0|\sigma_i^z|\Psi_0\rangle\langle\Psi_0|\sigma_j^z|\Psi_0\rangle \tag{3.16}$$

we see that we can write the correlation function as the susceptibilty of the magnetization at i with respect to a magnetic field η_j at j:

$$C(i,j) = \frac{\partial}{\partial\eta_j}\langle\Psi_0(\eta_j)|\sigma_i^z|\Psi_0(\eta_j)\rangle\Big|_{\eta_j\to 0}, \tag{3.17}$$

where $\Psi_0(\eta_j)$ is the ground state with the field at η_j. By perturbation theory this is

$$|\Psi_0(\eta_j)\rangle = |\Psi_0\rangle + \eta_j \sum_{n>0} |\Psi_n\rangle \frac{\langle\Psi_n|\sigma_j^z|\Psi_0\rangle}{E_n - E_0} + O(\eta_j^2) \tag{3.18}$$

and after rewriting the denominators as integrals over time we define $\sigma_j^z|\Psi_0\rangle = |\tilde{j}\rangle$ is the position ket of an excitation at j. Notice that, to lowest order in J/Γ, $|\tilde{j}\rangle$ is exactly the single x-spin flipped state $|j\rangle$ we introduced before. So, ignoring this difference, we can write

$$C(i,j) = i\int_0^\infty dt\ e^{iE_0t}\langle i|e^{-iHt}|j\rangle, \tag{3.19}$$

i.e. the spin–spin correlation function is the frequency component E_0 of the propagator of a particle created at j and detected at i:

$$C(i,j) = G(i,j|E_0). \tag{3.20}$$

Since the band is away from the ground state, the particles will be exponentially damped, irrespective of the fact that the states in the band are, according to this independent particle picture, delocalized.

So at the transition the picture is that of a finite density of particles which interact and annihilate in couples (like anyons) and whose gain in energy due to delocalization makes the vacuum state unstable. Remembering that the single particle eigenstates are excitations on top of the ground state of the paramagnetic phase this means that excitations are delocalized. Therefore there is no localization in the region of the parameters space containing the phase transition (in the particles mean field picture). The only possible source of localization for the excitations could from the disorder in the average magnetizations

$$\langle\sigma_i^z\rangle = \langle\Psi_0|\sigma_z^i|\Psi_0\rangle = m_i, \tag{3.21}$$

inside the glassy phase. The excitations would then feel an effective disorder due to the variance of the magnetization site-by-site: $W \propto \sqrt{q_{EA}}$. When this disorder exceeds a critical value function of both the hopping $\sim \Gamma$ and the effective interaction between the particles the particles would be localized, and this would be the onset of localization.

This point will be addressed in Chapter 4, where we anticipate that different numerical techniques will be needed as QMC numerics is not suited to study the dynamics of isolated quantum system.

We now move to the prediction of this mean field for the entanglement entropy of the ground state. Let us use the basis $|0\rangle, |i\rangle, |i,j\rangle, |i,j,k\rangle, \dots$. Then writing a generic state in this basis

$$|\Psi\rangle = c_0|0\rangle + \sum_{i=1}^N c_i|i\rangle + \sum_{i,j=1}^N c_{i,j}|i,j\rangle + \cdots\ . \tag{3.22}$$

The reduced density matrix $\rho_A = \mathrm{Tr}_B\,|\Psi\rangle\langle\Psi|$ can be broken in different blocks pertaining to different particle numbers

$$\rho_A = \rho_0 \oplus \rho_1 \oplus \cdots \oplus \rho_{N_A}, \tag{3.23}$$

where N_A is the number of sites in A, which is the maximum number of particles allowed in the region. Any Rényi entropy, including the entanglement entropy, will be written as a sum over the Rényi entropies of the different sectors:

$$\mathrm{Tr}\,\rho_A^2 = \sum_{n=1}^{N_A}\mathrm{Tr}\,\rho_n^2. \tag{3.24}$$

Now, the sector with n particles contains at most $\binom{N_A}{n}$ states, so it is clear that if we want to have

$$S^{(2)} = -\ln \operatorname{Tr} \rho_A^2 \propto N_A \tag{3.25}$$

as our numerics shows, we need to have $n \propto N_A$ so that at least one of the contributions $\operatorname{Tr} \rho_n^2$ is exponential in $N_A \propto N$. So, both the extensivity of the Rényi entropy and the correction to Γ_c point in the direction of a finite density of excitations in the ground state.

Summarizing this section, and refraining from doing numerology, there is a lesson to be learned from a mean-field theory of excitations on the x-polarized ground state. First of all, the starting point $\Gamma = \infty$ seems a good one, both qualitatively and quantitatively, in particular at large K. It is very instructive, in order to build an intuition of this glass transition, to consider this phenomenology complementary to the one obtained by starting from the classical z-spin glass at $\Gamma = 0$ and the dressing it with excitations. Then again, it seems that looking at different particle numbers sector n we get a better approximation going up with n so the transition occurs when a highly occupied state comes close to the ground state, which has only "virtual" particles.
This points in a direction which is completely different from that advocated in previous studies. The phase transition is dominated by quantum processes and probably has little signature of the complexity of the classical problem. Since it is reasonable to expect that the latter has to manifest itself at some point during the adiabatic evolution, it is conceivable that an inner region of the glassy phase is characterized by another phase transition. A natural candidate would be an MBL region, which is typically affected by exponentially small gaps, which would therefore be encountered by the adiabatic algorithm before the classical point. This research direction will be discussed in the next chapter.

Chapter 4

Localization Properties

In this chapter we study some dynamical properties of the transverse-field Ising spin glass model. In particular, we are interested in the existence of a localization/delocalization transition as the transverse field strength Γ is decreased from infinity. This was linked to poor performances of the Adiabatic Quantum Algorithm on random instances of NP-hard problems in the influential paper [67]. In Section 4.1 we review the phenomenon of Anderson localization and in Section 4.2 we see how this affects the quantum adiabatic algorithm according to Ref. [67]. In Section 4.3 we study the localization properties of the transverse-field Ising spin glass model using the Forward Scattering Approximation of the locator expansion. Numerical results are given in Section 4.4. In Section 4.6 we try to understand localization by studying the level statistics of the model.

4.1 Anderson Localization

In a seminal paper [63] published in 1958, Philip Anderson studied the suppression of transport in disordered materials. Anderson proposed a tight-binding model (nowaday called Anderson model) to describe spinless, non-interacting electrons hopping on a lattice and scattering off a random potential. Its Hamiltonian is given by

$$H_{\mathrm{AM}} = \sum_i \epsilon_i c_i^\dagger c_i - \sum_{\langle i,j \rangle} V_{ij} \left(c_i^\dagger c_j + c_j^\dagger c_i \right), \tag{4.1}$$

where the first term is the disordered local potential term with random energies ϵ_i independently and identically distributed with uniform probability in the interval $[-W/2, W/2]$. $W > 0$ is a constant that controls the strength of the disorder in the system. The coefficients V_{ij} are hopping coefficients between sites i and j, which we take to be equal to a constant $V > 0$ for all i, j. The sum in the hopping term is taken over the nearest neighbours in a d-dimensional hypercubic lattice $\mathbb{Z}^d$ (or a hypercube of finite size L^d with periodic boundary conditions, when a finite system is preferable).

In the limit of vanishing potential, $W = 0$, the hopping term is the discrete version of a kinetic term $-\nabla^2$. In the continuum limit, the eigenstates of the Hamiltonian are given by plane waves, and are associated to a continuous spectrum. In the position basis the eigenstates are extended, in the sense

that their wavefunction $\psi(x) = \langle x|\psi\rangle$ does not decay with the distance as one moves away from any fixed position x. This extended shape of the wavefunctions is conserved also in the discrete case over the lattice $\mathbb{Z}^d$ with position basis $\{|i\rangle \mid i \in \mathbb{Z}^d\}$. The dynamics of the system, generated by the kinetic term, is ballistic in the sense that for a generic initial state $|\psi\rangle$ one has that the mean square displacement goes like

$$\sum_{i\in\mathbb{Z}^d} |r(i)|^2|\psi(i,t)|^2 \sim t^2$$

for sufficiently large times t. Here $\psi(i,t) = \langle i|\psi(t)\rangle$, and $r(i)$ is the displacement of a lattice site i from a fixed site 0 in the lattice.

In the opposite limit $W \to \infty$ the hopping is completely suppressed and the eigenstates are delta functions $\langle j|\psi_i\rangle = \psi_i(j) = \delta_{ij}$ localized at site i: $H|i\rangle = \epsilon_i|i\rangle$. These eigenstates are obviously localized and correspond to a point spectrum (in both the continuous and the discrete case). The real-time evolution of the system satisfies

$$\sum_{i\in\mathbb{Z}^d} |r(i)|^2|\psi(i,t)|^2 \to \text{constant}$$

in the large time limit.

Given the very different behaviours of these two opposite limits, the expectation is that there will be a transition between them driven by the parameter V/W, which controls the relative strength of the disordered potential (that induces localization) and the hopping term (that delocalizes the system). Indeed what happens in general is that, by moving from the free theory in the direction of stronger disorder one finds that at some point in the middle of the delocalized spectrum of the systems an interval of energies appears whose associated energy eigenstates are localized. This interval then spreads until the entire spectrum is localized. For this reason, after having fixed the hopping constant V which sets the energy scale, a disorder-driven localization/delocalization transition is usually studied as a function of two quantities: the energy density ϵ of the eigenstates considered and the strength W of the disorder in the Hamiltonian. One usually finds that states at a fixed energy density ϵ change from delocalized to localized as the disorder strength W is increased, with an energy-dependent critial value $W_c(\epsilon)$ that marks the boundary between these two behaviours. The set of critical points $W_c(\epsilon)$ for different values of ϵ define the "mobility edge" of the system.

One can study the delocalization transition in the strong disorder regime by considering the hopping constant V as a small parameter and then expand perturbatively from the potential term. The first-order correction in V gives

$$|\psi_n\rangle = |n\rangle + \sum_{m\neq n} \frac{\langle m|H_{\text{hop}}|n\rangle}{\epsilon_n - \epsilon_m}|m\rangle,$$

where $H_{\text{hop}} = -V\sum_{\langle i,j\rangle} c_i^\dagger c_j + h.c.$ is the hopping term. Note that $\langle m|H_{\text{hop}}|n\rangle = 0$ unless the classical states $|m\rangle, |n\rangle$ are localized on adiacent sites in the lattice, so the sum can be taken over the nearest neighbours of the site n. If V is small, then these wavefunction amplitudes $|\langle m|\psi_n\rangle|$ for $m \neq n$ are small, unless for

some m it happens that $\epsilon_m \approx \epsilon_n$ and the $\langle m|H_{\text{hop}}|n\rangle/(\epsilon_n - \epsilon_m)$ term blows up. At second order we have

$$|\psi_n\rangle = |n\rangle + \sum_{m\neq n} \frac{\langle m|H_{\text{hop}}|n\rangle}{\epsilon_n - \epsilon_m}|m\rangle + \sum_{m,k\neq n} \frac{\langle m|H_{\text{hop}}|k\rangle\langle k|H_{\text{hop}}|n\rangle}{(\epsilon_n - \epsilon_m)(\epsilon_n - \epsilon_k)}|m\rangle$$

and the same argument applies, only this time we have to consider the energy differences between ϵ_n and site at distance two from n. One can see that at higher orders in the theory of perturbation one has to consider the probability of finding *resonances*, *i.e.* very small energy differences $\epsilon_n - \epsilon_m \approx 0$ arbitrarily far away from the initial site n.

Generally we have that as we move from weak to strong disorder, the wavefunctions change form a roughly uniformly spread, extended shape to a shape that decays exponentially as one moves away from a finite region of the lattice (see Fig. 4.1). In the first case we say that the eigenstate is *extended* while in the second case we say it is *localized*. We will give formal definitions of these terms in Section 4.3.

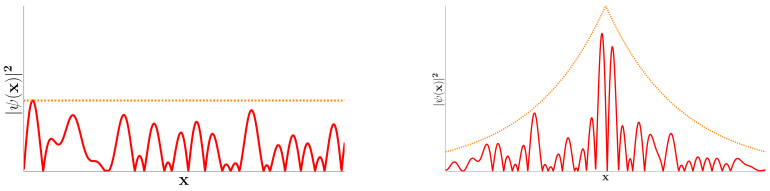


Figure 4.1: (Left) Extended state. The probability density $|\psi(x)|^2$ does not decay exponentially over the space coordinate x. (Right) Localized state. The probability density $|\psi(x)|^2$ is contained under an exponentially-decaying envelope centered at a point called the *localization center*. The coefficient ξ of the exponential decay is called the *localization length*.

At this point we have three apparently independent properties of the system that should undergo a transition driven by the parameter V/W:

1. the dynamics changes from ballistic to essentially freezed
2. the spectrum changes from a continuous to a point spectrum (in the continuous limit)
3. the energy eigenstates change from extended to localized

Important results in mathematical physics show that these properties are *not* independent, but there are ways to relate them.

The connection between spectral properties and dynamics can be formulated using the RAGE theorem[1] [68, 69, 70], which we briefly state. In order to study the real-time dynamics of a quantum system defined over a graph $G = (V, E)$ (such as the Anderson model of Eq. (4.1)) and define the Hilbert space $\mathcal{H} =$

[1] from the name of its authors, Ruelle, Amrein, Georgescu and Enss

$\ell^2(G)$ to be the square-summable sequences of complex numbers indicized by the set of vertices V

$$\ell^2(G) = \left\{ \psi : V \to \mathbb{C} \;\middle|\; \sum_{v \in V} |\psi(v)|^2 < \infty \right\}$$

For a fixed Hamiltonian H defined on this Hilbert space, and an initial state $|\psi(0)\rangle \equiv |\psi\rangle$, let $|\psi(t)\rangle = e^{-iHt}|\psi\rangle$ be the time-evolved sate. The idea we want to capture is whether a given initial state will be spread by the dynamics over arbitrarily large regions of $\mathcal{H}$ or whether it will remain contained in a finite region. Thus, for a given subset of vertices $W \subseteq V$ we indicate with $\Pr[W, t]$ the probability that the state $|\psi(t)\rangle$ is inside of W

$$\Pr\Big[W, t\Big] = \sum_{w \in W} |\langle w|\psi(t)\rangle|^2$$

It turns out that the spreading of a state generated by the dynamics is dependent on some spectral characteristics of the Hamiltonian H: given a self-adjoint operator H and a state $|\psi\rangle \in \mathcal{H}$, then the spectral theorem [71] allows us to write the expectation values with respect to the state $|\psi\rangle$ of any operator $f(H)$, where f is a bounded and limited functions $f : \mathbb{R} \to \mathbb{C}$, as the integral of the energy function $f(E)$ with respect to a measure $\mu_\psi(E)$

$$\langle\psi|f(H)|\psi\rangle = \int_{-\infty}^{\infty} f(E)\, \mathrm{d}\mu_\psi(E).$$

This measure $\mu_\psi(E)$ admits a Lebesgue decomposition

$$\mu_\psi = \mu_\psi^{\mathrm{ac}} + \mu_\psi^{\mathrm{sc}} + \mu_\psi^{\mathrm{pp}}$$

into an *absolutely continuous* component (with respect to the Lebesgue measure over $\mathbb{R}$), a *singular continuous* component and a *pure point* component. This decomposition induces a direct-sum decomposition of the Hilbert space into orthogonal subspaces

$$H = \mathcal{H}^{\mathrm{ac}} \oplus \mathcal{H}^{\mathrm{sc}} \oplus \mathcal{H}^{\mathrm{pp}},$$

where a state $|\psi\rangle$ belongs to $\mathcal{H}^{\mathrm{ac}}$ if and only if its spectral measure μ_ψ coincides with its absolutely continuous component, and analogously for the singular continuous and pure point components. Then the RAGE theorem describes the evolution $e^{-iHt}|\psi\rangle$ of an initial state $|\psi\rangle$ depending on which of these subspaces the state belongs to.

Theorem 3 (RAGE Theorem, [72])**.** *Let H be a Hamiltonian defined on the Hilbert space $\mathcal{H} = \ell^2(G)$, where $G = (V, E)$ is a graph. Then the following facts hold.*

- *If the initial state $|\psi\rangle$ belongs to the absolutely continuous subspace $\mathcal{H}^{\mathrm{ac}}$, then for any ball of fixed radius r centered on any vertex $v \in V$*

$$\lim_{t \to \infty} \Pr\Big[B(v, r), t\Big] = 0.$$

  This means that the systems leaves any finite region.

- *If the initial state $|\psi\rangle$ belongs to the pure point subspace $\mathcal{H}^{\mathrm{pp}}$, then*

$$\lim_{r\to\infty} \sup_t \Pr\Big[B(0,r)^c, t\Big] = 0$$

This means that the system is confined for all time to a finite region.

Of course these statements are meaningful only for systems defined on infinitely-large graphs, otherwise the limits $r \to \infty$ become pathological. By the RAGE theorem, dynamical localization is related to spectral localization: a Hamiltonian H is localized in the energy interval $[E, E + \Delta E]$ if the subset of its spectrum $\mathrm{Sp}(H)$ contained in $[E, E + \Delta E]$ is only pure point.

4.2 Anderson Localization and AQC

In the previous section we have briefly described some features of the theory of Anderson localization but up to now these seems unrelated to the underlying topic of this thesis, which is quantum annealing. The following is an intuitive explanation of the reason why localization and the suppression of transport might play a role in adiabatic quantum computation. A possible picture of the adiabatic algorithm is a process defined by the time-dependent energy landscape described by the Hamiltonian $H(t)$ used in the annealing process. If we consider the quantum annealing of the transverse-field Ising spin glass Hamiltonian of Eq. (3.1) then we have that at large values of Γ the energy landscape of the system will be dominated by the transverse-field term, with a single potential well with a non-degenerate ground state. As Γ is decreased below the the critical value Γ_c that marks the boundary of the glassy phase, new local minima will start to appear and their relative heights and positions will get shifted as Γ changes.

At each avoided crossing encountered in the adiabatic path, the energy of ground state falls below the energy of a metastable state, so that this metastable state becomes the new ground state. In order to follow the adiabatic path, the system then has to tunnel from the old ground state to the new, in a process similar to an escape from a metastable state. The intuitive idea is that if we are in a regime where tunnelling is suppressed exponentially, this will require an exponentially long time, thus rendering the adiabatic algorithm inefficient.

More rigorously, the connection between Anderson localization and the Adiabatic Quantum Algorithm was established in a paper by Altshuler *et al.* [67] (see also the subsequent papers [73, 74]), where it was shown that by writing the Hamiltonian of a typical quantum annealing protocol $H = H_0 + \Gamma \sum_i \sigma_i^x$ in the σ^z (*i.e.* "classical") basis, one gets

$$H = \sum_s \epsilon_s |s\rangle\langle s| - \Gamma \sum_{s,s':\mathrm{dist}(s,s')=1} \Big(|s\rangle\langle s'| + |s'\rangle\langle s|\Big) \tag{4.2}$$

where $\epsilon_s \equiv f(s)$ are the classical energies of the COP Hamiltonian $H_0 = \sum_s f(s)|s\rangle\langle s|$, and the sum in the second term is taken over pairs s, s' that differ only by exactly one spin flip, as the matrix element of the transverse-field term satisfies

$$\langle s| \sum_i \sigma_i^z |s'\rangle = \begin{cases} 1 & \text{if } |s\rangle = \sigma_j^z |s'\rangle \text{ for some } j \\ 0 & \text{otherwise.} \end{cases}$$

One can see that the first term in Eq. (4.2) can be interpreted as the analogue of the local potential of the Anderson model while the second term defines the hopping of a particle on the N-dimensional Boolean hypercube $\mathbb{B}_N = \{\pm 1\}^N$. Notice in this case that the local energies ϵ_s are no longer independent but correlated, as they are the energy levels of a system with complicated interactions (*i.e.* , the Hamiltonian that encodes constraint satisfaction problem). In a typical setup of the Adiabatic Quantum Algorithm, one starts with a large value of Γ that is slowly decreased to zero. In the small Γ regime, the disorderd potential will dominate on the hopping term and one might expect that localization will set in, as it happens in the Anderson model.

In [67], the classical Hamiltonian H_0 is not a spin glass model but an encoding of the decision problem known as 3-Exact-Cover, described in Section 1.2. Nevertheless, most of the details apply for any reasonable constraint satisfaction problem.

The classical Hamiltonian H_0 that codifies an instance $f = \sum_i f_i$ of 3-Exact-Cover instances can be chosen of the form

$$H_0 = M\mathbb{I} - \frac{1}{2}\sum_{i=1}^{N} B_i \sigma_i^z + \frac{1}{4}\sum_{i,j=1}^{N} J_{ij}\sigma_i^z\sigma_j^z$$

where M is the total number of constraints, B_i is the number of constraints where the variable X_i appears and J_{ij} is the number of constraints where *both* of the variables X_i, X_j appear. Satisfying solutions x to the instance are associated to eigenstates $|x\rangle$ of zero energy, $H_0|x\rangle = 0$.

The main idea behind the Altshuler *et al.* paper is the following. Suppose we have a 3-Exact-Cover instance with f two solutions x, y, *i.e.* $f(x) = f(y) = 0$. In the Hamiltonian H_0 of the combinatorial optimization problem, these will correspond to the classical, localized states $|x\rangle$ and $|y\rangle$ with energies $E_x = E_y = 0$. By turning on the transverse field, their degeneracy will be split and the energy of the two solutions will be shifted by different amounts. If we increase Γ, the energy difference between the state $|\psi_x(\Gamma)\rangle$ and $|\psi_y(\Gamma)\rangle$ will eventually exceed the $O(1)$ energy contribution that a single constraint can contribute to the total energy of the system. If this is the case, then a modified instance with an additional constraint satisfied only by one of the two solutions x, y will introduce an avoided crossing in the annealing path (see Fig. 4.2). The authors first compute the position Γ^* of such an avoided crossing, then the width of its minimal gap. Since Γ^* is found to be very small, it follows that if the solutions $|x\rangle, |y\rangle$ are very different from each other (say, $O(N)$-far in the Hamming distance) then the perturbed states $|\psi_x(\Gamma^*)\rangle$ and $|\psi_y(\Gamma^*)\rangle$ will also be very different simply because the perturbation is not strong enough to affect the states $|x\rangle, |y\rangle$ significantly. Then it will be hard for the system to tunnel from the old ground state $|\psi_x(\Gamma)\rangle$ to the new one $|\psi_y(\Gamma)\rangle$. This translates into a small energy gap at the avoided crossing, which is indeed found to close faster than exponentially with the system size.

In the following we review the main points in the derivation. The self-energy operator for the Hamiltonian $H(\Gamma)$ is defined as

$$\Sigma(E, \Gamma) \equiv G_0(E)^{-1} - G(E, \Gamma)^{-1}$$

where $G_0(E) = (E - H_0)^{-1}$ is the Green's function of the classical Hamiltonian

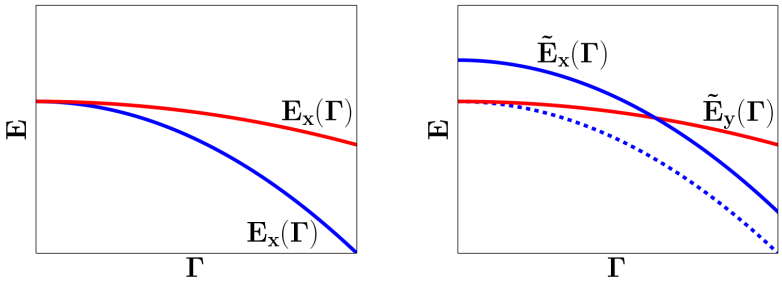


Figure 4.2: (Left) the two solutions x, y satisfy $E_x(\Gamma) = E_y(\Gamma) = 0$ at $\Gamma = 0$. However, if the transverse field is turned on, their two energy levels will split. (Right) By adding a constraint to the 3-Exact-Cover instance that is satisfied by solution y but violated by x, the shift in the energy level $E_x(\Gamma)$ will create an avoided crossing at a point Γ^* where the energy difference $E_x(\Gamma) - E_y(\Gamma)$ between the two levels is large enough. The requirement that one needs to be sure that $\Delta E_{xy}(\Gamma)$ is larger than the $4 = O(1)$ energy contribution of a single constraint is necessary to avoid the possibility that the $\tilde{E}_x(\Gamma)$ level of the new instance lies above $\tilde{E}_y(\Gamma)$ for all Γ.

H_0 and $G(E,\Gamma) = (E - H(\Gamma))^{-1}$ is the Green's function of the perturbed Hamiltonian $H(\Gamma)$. By taking the diagonal part of the self-energy with respect to a classical basis vector $|x\rangle$ we define the self-energy of x as

$$\Sigma_x(E,\Gamma) \equiv \langle x|\Sigma(E,\Gamma)|x\rangle.$$

Note that this quantity can be expanded in series of Γ as follows

$$\Sigma_x(E,\Gamma) = \sum_{q=1}^{\infty} \Gamma^q \sum_{y_1,\ldots,y^{q-1}} \frac{\langle x|V|y^1\rangle\langle y^1|V|y^2\rangle\cdots\langle y^{q-1}|V|x\rangle}{(E - E_{y^1})\cdots(E - E_{y^{q-1}})} \tag{4.3}$$

where none of the classical state $y_1, \ldots, y^{q-1}$ is equal to x and $V = -\sum_i \sigma_i^x$ is the transverse-field perturbative term. The diagonal part of the perturbed Green's function is related to the diagonal part of the self-energy operator through the identity

$$G_x(E,\Gamma) = \frac{1}{E - E_x - \Sigma_x(E,\Gamma)}. \tag{4.4}$$

Since the poles $\{E_i\}$ of the Green's function $G(E,\Gamma)$ correspond to the eigenvalues of the Hamiltonian $H(\Gamma)$, then we have that Eq. (4.4) implies that the energy levels $E_x(\Gamma)$ must satisfy the equation $E - E_x - \Sigma_x(E,\Gamma) = 0$, which is equivalent to

$$E = E_x + \Sigma_x(E,\Gamma). \tag{4.5}$$

This is a self-consistent equation that can be solved iteratively through repeated

substitutions of $E \to E_x + \Sigma_x(E, \Gamma)$, obtaining

$$\begin{aligned} E &= E_x + \Sigma_x(E, \Gamma) \\ &= E_x + \Sigma_x(E_x + \Sigma_x(E, \Gamma), \Gamma) \\ &= E_x + \Sigma_x(E_x + \Sigma_x(E_x + \Sigma_x(E, \Gamma), \Gamma), \Gamma) \\ &= \cdots \end{aligned}$$

It is not at all obvious what a solution should look like in the infinite limit of this iterative process. However, we can use the expansion of Eq. (4.3) in order to obtain a perturbative expansion for the energy level $E_x(\Gamma)$

$$E_x(\Gamma) = \sum_{k=0}^{\infty} \Gamma^k E_x^{(k)}$$

using the fact that for any fixed order k, the energy contribution $E_x^{(k)}$ can be computed from an equation obtained from Eq. (4.5) by appying only a finite number of self-consistent substitutions.

Given two solutions x, y then define $\Delta E_{xy}(\Gamma) \equiv E_x(\Gamma) - E_y(\Gamma)$ as the energy gap between them. Then one can plug in the expansion for $E_x(\Gamma), E_y(\Gamma)$ and obtain a perturbative expansion for the gap

$$\Delta E_{xy}(\Gamma) = \sum_{q=0}^{\infty} \Gamma^q \Delta E_{xy}^{(q)},$$

where $\Delta E_{xy}^{(q)} = E_x^{(q)} - E_y^{(q)}$. For the Exact-Cover problem, the first non-zero correction is $\Delta E_{xy}^{(4)}$, as combinatorial properties entail that $E_x^{(q)} = 0$ for all odd q and that the correction to the energy $E_x^{(2)}$ is independent of the choice of x. Then the leading order of $\Delta E_{xy}(\Gamma)$ is given by $\Gamma^4 \Delta E_{xy}^{(4)}$. Moreover, the term $\Delta E_{xy}^{(4)}$ is a sum of $O(N)$ random variables of zero mean, so we expect its variance to be of order $O(N)$

$$\left\langle (E_{xy}^{(4)})^2 \right\rangle - \left\langle E_{xy}^{(4)} \right\rangle^2 = \left\langle (E_{xy}^{(4)})^2 \right\rangle \approx C^{(4)} N,$$

which in turn implies that the ϵ-th percentile is of order $C_\epsilon^{(4)} N = O(N)$. This means that $\Pr[(E_{xy}^{(4)})^2 \geq C_\epsilon^{(4)} N] \approx 1 - \frac{\epsilon}{100}$. So, with probability $1 - \frac{\epsilon}{100}$ we have

$$\Delta E_{xy}(\Gamma) = \Gamma^4 E_{xy}^{(4)} \quad \geq \quad \Gamma^4 (C_\epsilon^{(4)} N)^{1/2}.$$

Now, since the energy contribution of a single constraint is 4, we need to find out how strongly we must perturb the classical Hamiltonian $H(\Gamma = 0)$ to find a gap that exceeds 4. In order to have $\Gamma^4 (C_\epsilon^{(4)} N)^{1/2} > 4$ we have to take $\Gamma > \Gamma^* = \sqrt{2} (C_\epsilon^{(4)} N)^{-1/8}$. If this is the case then the introduction of a new constraint that is satisfied by the assignment x but not by y introduces an avoided crossing at the point Γ^*, as shown in Fig. 4.2. Since at the avoided crossing the ground state of the system changes from the wavefunction adiabatically connected to $|x\rangle$ to the wavefunction adiabatically connected to $|y\rangle$ (similarly to what happens in a Landau-Zener anticrossing, see Section 1.3.2), then in order to follow the

adiabatic path the system has to tunnel from $|\psi_x(\Gamma)\rangle$ to $|\psi_y(\Gamma)\rangle$. The tunnelling amplitude can be computed as

$$A_{xy}(\Gamma) = \langle x|\psi_y(\Gamma)\rangle.$$

The authors perturbatively compute $A_{xy}(\Gamma^*)$, which gives the gap g at the point Γ^* of the avoided crossing, finding that is scales with the system size N as

$$g(N) = 2^{-O(N \log N)}.$$

This is a superexponentially-closing gap, which means that the Adiabatic Quantum Algorithm on a random instance performs worse than the classical solvers (which at worse require exponential time). These results, however, are based on the assumptions that the perturbation theory starting from the classical $H(\Gamma = 0)$ Hamiltonian is convergent. We have seen that a system such as that of Eq. (3.1) is localized at $\Gamma = 0$ and delocalized at $\Gamma = \infty$. If the localization/delocalization transition happens at a critical value of Γ_c smaller than Γ^* then the perturbation expansion will diverge and all conclusions for values of Γ larger than Γ_c are invalid.

The publication of [67] stirred up reactions in the scientific community. In [75] it was argued that the argument presented in [67] only shows that a randomly chosen energy level belonging to the ground state band and a randomly chosen level in the first band will have an avoided crossing with the described properties. These two levels need not evolve to the ground state and the first excited state of the system when the perturbation is turned on unless their respective bands are non-degenerate, which is extremely unlikely as the spectra of constrain satisfaction problems are typically expected to have exponential degeneracy. By considering an exponential density of states and focussing on the ground state and the first excited state only, the authors obtain a crossing at a distance $O(1)$ away from $\Gamma = 0$, which might be outside of the convergence radius of the perturbation theory. Others have proposed ways to go around the avoided crossing, at least in some cases [76, 77]. While controversial, the [67] paper is still one of the main stepping stones in the understanding of the physics behind the Adiabatic Quantum Algorithm.

4.3 Localization in the transverse-field Ising spin glass model

In light of the results described in the previous section, our goal for the rest of the chapter is to study the localization properties of the transverse-field Ising spin glass model $H(\Gamma)$ on the 3-regular random graph.

$$H(\Gamma) = -\sum_{\langle ij \rangle} J_{ij}\sigma_i^z\sigma_j^z - \Gamma\sum_i \sigma_i^x \tag{4.6}$$

As usual in our setup we keep the strength of the disordered interactions fixed $W \equiv |J_{ij}| = 1$ so the relative strength of the disorder with respect to the ordering term $-\Gamma\sum_i \sigma_i^x$ is controlled by the parameter $W/\Gamma = 1/\Gamma$. The mobility edge will consequently be defined by the critical values $\Gamma_c(\epsilon) \equiv 1/W_c(\epsilon)$.

In particular, we want to understand if, for a given value of Γ and a given energy density interval $\mathcal{E} = [\epsilon, \epsilon + \Delta\epsilon]$, the eigenstates $|\psi(\Gamma)\rangle$ of $H(\Gamma)$ with energy density in $\mathcal{E}$ are exponentially-decaying in the σ^z basis or not.

Let us better define the kind of localization we will study. First, we fix a value ϵ for the energy density and Γ for the strength of the transverse field. For a given realization of disorder of the Hamiltonian $H(\Gamma)$ of Eq. (4.6) defined on a system of size N and a energy eigenstate $|\psi\rangle$ with localization center $|v\rangle$ and energy density ϵ we define

$$\psi_r \equiv \max\Big\{|\langle w|\psi\rangle| : w \in \{\pm 1\}^N, \mathrm{dist}(v,w) = r\Big\} \qquad r = 1, 2, \ldots \tag{4.7}$$

where the distance is taken over the Boolean hypercube[2]. Next, for fixed $N, r \in \mathbb{N}$ and real numbers $C, \xi > 0$ we define the quantity $P(N, r, \xi)$ as the probability (taken over all disorder realizations of $H(\Gamma)$ of size N and energy eigenstates of energy density ϵ) that the random variable $|\psi_r|^2$ satisfies $|\psi_r|^2 \leq Ce^{-r/\xi}$

$$P(N, r, \xi) \equiv Pr\Big(|\psi_r|^2 \leq Ce^{-r/\xi}\Big). \tag{4.8}$$

This means that the probability distribution $p(x) = |\psi(x)|^2$ is under an exponentially-decaying envelope, except possibly for a region of finite radius, for all disorder realizations except for a fraction that vanishes in the thermodynamic limit. If no such exponential bound can be established, then the state is delocalized. What usually happens in this case, however, is something stronger. For all $\xi > 0$ the probability of not being under any such exponential envelope goes to one as the system size is increased

$$1 - P(N, r, \xi) \equiv Pr\Big(|\psi_r|^2 > Ce^{-r/\xi}\Big).$$

This suggests the following definitions.

Definition 5 (Localized/Extended state)**.** *We say that the disordered Hamiltonian $H(\Gamma)$ is* localized *at energy density ϵ if for all there exist a real number $\xi > 0$ such that*

$$\lim_{r\to\infty} \lim_{N\to\infty} P(N, r, \xi) = 1.$$

We say that the $H(\Gamma)$ is extended *at energy density ϵ if for all $\xi > 0$ we have*

$$\lim_{r\to\infty} \lim_{N\to\infty} P(N, r, \xi) = 0.$$

Notice that in place of Eq. (4.8) we can write equivalently

$$P(N, r, \xi) = P\Big(\frac{\ln|\psi_r|^2}{2r} \leq -\frac{1}{2\xi} + O\Big(\frac{1}{r}\Big)\Big),$$

[2]Note that in order to compute ψ_r from the amplitudes $|\langle w|\psi\rangle|$ one needs to know (or find out) the localization center of the state $|\psi\rangle$. Since $|\psi\rangle$ is an eigenstate of the perturbed Hamiltonian $H(\Gamma)$ then in the limit $\Gamma \to 0$ it will be connected to an unperturbed eigenstate $|a\rangle$, which is a Kronecker delta in the unpertubed energy eigenbasis (*i.e.* it exhibits an extreme form of localization). If small perturbations do not destroy the localization, then the perturbed eigenstate $|\psi\rangle = |\psi_a(\Gamma)\rangle$ will still be localized around $|a\rangle$. Therefore in the following Sections we will start from the working hypothesis that if $|\psi_a(\Gamma)\rangle$ is localized, then $|a\rangle$ is its localization center.

so we can study the distribution of the values of the random variable $Z_r \equiv \ln|\psi_r|/r$. It was observed numerically in Ref. [78] (see Fig. 4.3) that if the system is localized then in the limit $N \to \infty$ the random variable Z_N is peaked around the value

$$Z_N \equiv \frac{\ln|\psi_N|}{N} \to \ln(\Gamma/\Gamma_c) \qquad \text{as } N \to \infty, \tag{4.9}$$

which gives the value of $\Gamma_c(\epsilon)$. In practice one can calculate ψ with $\Gamma = 1$ and get Γ_c.

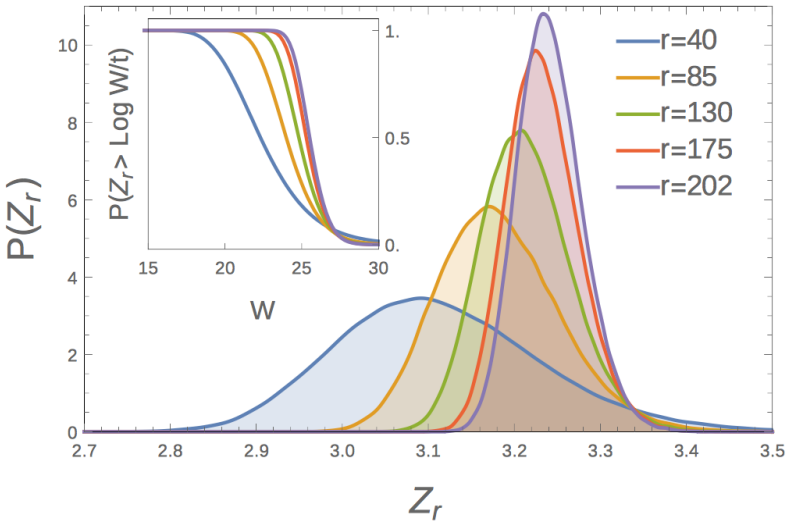


Figure 4.3: taken from Ref. [78]. Probability density of the variable Z_r, for different r, computed for the Anderson model on the cubic lattice $\mathbb{Z}^3$. Note the concentration of the distribution of Z_r on the point $\log(W_c/t)$ (equivalent to $\log(\Gamma/\Gamma_c)$ in our notation) as r is increased.

4.3.1 The Forward Scattering Approximation

We have seen how one can statistically study the localization properties of disordered systems by sampling wavefunction values $|\psi(x)|^2$. Of course this is not always a simple task, as in general one needs to diagonalize a Hamiltonian whose size grows exponentially with the size of the system. However, in a setup such as the one defined by a quantum annealing protocol, where we are interested in studying the localization of the energy eigenstates of a perturbed Hamiltonian $H = H(\Gamma)$

$$H(\Gamma) = H_0 + \Gamma V$$

in the basis of the (simpler) unperturbed energy eigenstates $|a\rangle$ then one can use a technique known as the *forward scattering approximation*.

Now and throughout this section we use Latin letters $a, b, c, \ldots, x, y, z$ to label energies and eigenstates of the unperturbed Hamiltonian H_0, such as $|x\rangle$

and E_x, while the perturbed eigenstates and energies will be indicated as $|\psi_x(\Gamma)\rangle$ and $E_x(\Gamma)$. It is intended that in the limit $\Gamma \to 0$ we have $|\psi_x(\Gamma)\rangle \to |x\rangle$ and $E_x(\Gamma) \to E_x$.

First, note that we can extract the values $\psi_a(x,\Gamma) = \langle x|\psi_a(\Gamma)\rangle$ of the energy eigenfunctions $|\psi_a(\Gamma)\rangle$ by writing the resolvent $G(E,\Gamma)$ of the perturbed Hamiltonian $H(\Gamma)$ in the unperturbed energy eigenbasis

$$G_{ab}(E,\Gamma) \equiv \langle a|G(E,\Gamma)|b\rangle = \langle a|\frac{1}{E-H(\Gamma)}|b\rangle = \sum_c \frac{1}{E-E_c(\Gamma)}\langle a|\psi_c(\Gamma)\rangle\langle\psi_c(\Gamma)|b\rangle$$

Assuming that the perturbed energy $E_a(\Gamma)$ is non-degenerate, we can write

$$\Big(E-E_a(\Gamma)\Big)G_{ab}(E,\Gamma) = \langle a|\psi_a(\Gamma)\rangle\langle\psi_a(\Gamma)|b\rangle + \sum_{a'\neq a}\frac{E-E_a(\Gamma)}{E-E_{a'}(\Gamma)}\langle a|\psi_{a'}(\Gamma)\rangle\langle\psi_{a'}(\Gamma)|b\rangle$$

so by taking the limit $E \to E_a(\Gamma)$ we have

$$\lim_{E\to E_a(\Gamma)}\Big(E-E_a(\Gamma)\Big)G_{ab}(E,\Gamma) = \langle a|\psi_a(\Gamma)\rangle\langle\psi_a(\Gamma)|b\rangle \tag{4.10}$$

Then we can extract $|\psi_a(b,\Gamma)|^2$ by taking $a=b$. We assume we do not know the perturbed resolvent G but we do know the unperturbed resolvent G_0. Then we can use the standard perturbative expansion in Γ

$$G = G_0 + \Gamma G_0 V G_0 + \Gamma^2 G_0 V G_0 V G_0 + \Gamma^3 G_0 V G_0 V G_0 V G_0 + \cdots$$

and by inserting resolutions of the identity in the unperturbed energy eigenbasis ($\mathbb{I} = \sum_k |k\rangle\langle k|$) we write the Green's function $G_{ab}(E,\Gamma)$ as a series whose n-th order term is of the form

$$\Gamma^n \sum_{k_1,\ldots,k_{n-1}} \frac{1}{E-E_a}\frac{1}{E-E_{k_1}}\cdots\frac{1}{E-E_{k_{n-1}}}\frac{1}{E-E_b}\langle a|V|k_1\rangle\langle k_1|V|k_2\rangle\cdots\langle k_{n-1}|V|b\rangle. \tag{4.11}$$

In the case of Hamiltonian (4.6) we have $V = -\sum_i \sigma_i^x$ and the kets $|k\rangle$ can be taken to be products states of the local σ^z eigenstates, $|\uparrow\rangle$ and $|\downarrow\rangle$, so $\langle k|V|k'\rangle = -1$ if the states $|k\rangle$ and $|k'\rangle$ are linked by exactly one spin flip (*i.e.* $|k'\rangle = \sigma_i^x|k\rangle$ for some position i) and $\langle k|V|k'\rangle = 0$ otherwise. This means that not all terms in the sum over $k_1,\ldots,k_{n-1}$ will contribute, but only those that define a path in the Boolean hypercube. Thus the n-th order term (4.11) can be rewritten as

$$(-\Gamma)^n \sum_{\substack{p\in\text{paths}(a,b)\\|p|=n}} \frac{1}{E-E_a}\frac{1}{E-E_{k_1}}\cdots\frac{1}{E-E_{k_{n-1}}}\frac{1}{E-E_b}$$

Assuming the series converges absolutely in some region, we can write the Green's function as

$$G_{ab}(E,\Gamma) = \sum_{p\in\text{paths}(a,b)}(-\Gamma)^{|p|}\Big(\prod_{k\in p}\frac{1}{E-E_k}\Big) \tag{4.12}$$

Note that this path representation of the Green's function is a sum of infinitely many terms, even if the system is finite-dimensional. This is because the

r.h.s of Eq. (4.12) contains *repeating paths*: paths that visit the same vertices more than once (actually, an arbitrary number of times). Moreover, the meromorphic structure of the Green's function is not conserved at any finite order (*e.g.* the expansion (4.12) does not have $E_a(\Gamma)$ as simple poles) so we cannot extract the amplitudes of the perturbed eigenstates from its residues. To solve these problems one needs to re-sum the series. For each site i, the sum of all loops around i is used to "renormalize" the weight of the paths. One defines the self-energy term $\Sigma_a(E,\Gamma)$ through the identity

$$G_a(E,\Gamma) \equiv \frac{1}{E - E_a - \Sigma_a(E,\Gamma)}, \tag{4.13}$$

where $G_a(E,\Gamma) \equiv G_{aa}(E,\Gamma)$. Remember that $\Sigma_a(E,\Gamma)$ is equal to the sum of the amplitudes of all the closed paths in which site a appears only as starting and ending point (see Eq. (4.3)). Note that $\Sigma_a(E,\Gamma)$ has a zero first-order term:

$$\Sigma_a(E,\Gamma) = \sum_{j:\langle a,j\rangle} \frac{\Gamma^2}{E - E_j} + O(\Gamma^3), \tag{4.14}$$

where the sum is over the sites j adiacent to a in the Boolean hypercube.

After this is done, one obtains a series over the set paths$^*(a,b)$ of *simple (non-repeating) paths* connecting the sites a and b:

$$G_{ba}(E,\Gamma) = \frac{1}{E - E_a - \Sigma_a(E,\Gamma)} \sum_{p\in \text{paths}^*(a,b)} (-\Gamma)^{|p|} \prod_{i\in p} \frac{1}{E - E_i - \Sigma_i^{(p)}(E,\Gamma)}. \tag{4.15}$$

The path-dependent term $\Sigma_i^{(p)}(E,\Gamma)$ is a self-energy term defined on the reduced graph where the sites $(a, 1, \cdots, i-1)$ previously visited by the path p have been removed.
Now using Eqs. (4.10) and (4.15):

$$\begin{aligned}
\langle b|\psi_a(\Gamma)\rangle &= \frac{1}{\langle \psi_a(\Gamma)|a\rangle} \lim_{E\to E_a(\Gamma)} \Big(E - E_a(\Gamma)\Big) G_{ba}(E,\Gamma) \\
&= \frac{1}{\langle \psi_a(\Gamma)|a\rangle} \lim_{E\to E_a(\Gamma)} \frac{E - E_a(\Gamma)}{E - E_a - \Sigma_a(E,\Gamma)} \sum_{p\in \text{paths}^*(a,b)} (-\Gamma)^{|p|} \prod_{i\in p} \frac{1}{E - E_i - \Sigma_i^{(p)}(E,\Gamma)} \\
&= \frac{\langle \psi_a(\Gamma)|a\rangle\langle a|\psi_a(\Gamma)\rangle}{\langle \psi_a(\Gamma)|a\rangle} \lim_{E\to E_a(\Gamma)} \sum_{p\in \text{paths}^*(a,b)} (-\Gamma)^{|p|} \prod_{i\in p} \frac{1}{E - E_i - \Sigma_i^{(p)}(E,\Gamma)} \\
&= \langle a|\psi_a(\Gamma)\rangle \sum_{p\in \text{paths}^*(a,b)} (-\Gamma)^{|p|} \prod_{i\in p} \frac{1}{E_\alpha - E_i - \Sigma_i^{(p)}(E_a(\Gamma),\Gamma)}.
\end{aligned}$$

Now we keep the leading-order contribution in Γ to $\langle b|\psi_a(\Gamma)\rangle$. Taking into account that the self-energies are $\Sigma_a(E,\Gamma) = O(\Gamma^2)$, we have to lowest order

$$\begin{aligned}
E_a(\Gamma) &\to E_a \\
\langle a|\psi_a(\Gamma)\rangle &\to 1 \\
\Sigma_i(E,\Gamma) &\to 0,
\end{aligned}$$

which gives

$$\langle b|\psi_a(\Gamma)\rangle = \sum_{p\in \mathrm{spaths}(a,b)} (-\Gamma)^{|p|} \prod_{i\in p} \frac{1}{E_a - E_i}, \tag{4.16}$$

where the set $\mathrm{spaths}(a,b) \subset \mathrm{paths}^*(a,b)$ contains the shortest paths from a to b.

While strictly speaking the forward-scattering approximation is an uncontrolled approximation (no error bounds are known), it has been used since Anderson's seminal work [63] and checked numerically in many cases (*e.g.* [78]). It was found to be convergent in the localized phase and divergent[3] in the delocalized phase.

Notice that the sum over paths of Eq. (4.16) can be computed numerically using a transfer matrix technique, in which case one computes iteratively the vector $|x_t\rangle \in \ell^2(\mathbb{B})$ defined by

$$|x_t\rangle = D_a \cdot A \cdot |x_{t-1}\rangle \tag{4.17}$$

with

$$\langle i|D_a|j\rangle = \delta_{ij} \frac{1}{E_a - E_i}, \tag{4.18}$$

$$\langle i|A|j\rangle = \langle i|V|j\rangle, \tag{4.19}$$

$$|x_0\rangle = |a\rangle \tag{4.20}$$

and

$$\langle b|\psi_a(\Gamma)\rangle = \langle b|x_N\rangle \tag{4.21}$$

where N is the distance between a and b in the Boolean hypercube. One can decrease memory requirements by noting that the vector v is very sparse during most of the computation, so at each step most entries of the transfer matrix $T = D_a A$ are irrelevant. This is because repeated applications of T to the initial state define a diffusion process on the Boolean hypercube where at each step t one needs to propagate only the amplitudes of the vertices exactly at distance t from the initial vertex. This means that in practice one does not need to store in memory the entire transfer matrix T, but instead a new transfer matrix T_t is defined at each step that only propagates amplitudes from vertices actually relevant for that single step of propagation. This requires storing only $(N-t)\binom{N}{t}$ non-zero entries instead of $N2^N$ of the full transfer matrix $T = D_a A$. The vector $|x_t\rangle$ need to store only $\binom{N}{t}$ entries. The full details of this method and a study of its time and memory requirements are contained in Appendix A.1.

4.4 Numerical results

In this section we apply the previously-described methods to the transverse-field Ising spin glass Hamiltonian. Here we immediately face an issue: the computation of the forward approximation is obstructed by the fact that the spin glass term $H_0 = -\sum_i J_{ij}\sigma_i^z\sigma_j^z$ has highly degenerate energy levels. This gives rise to diverging terms in Eq. (4.16) when $E_a = E_i$. In order to avoid

[3] in the thermodynamic limit. For any finite N Eq. (4.16) only has finitely-many terms.

this problem we add a weak and random longitudinal field term H_{long} to the Hamiltonian H_0:

$$H_{\text{long}} = -\sum_i h_i \sigma_i^z$$

where each h_i is distributed uniformly in $(-h, h)$ with $h = 0.001$ (this has to be $\ll 1/N$ but $\gg e^{-aN}$). This has the effect of splitting the degeneracies while introducing only a negligible effect in the energies of the configurations (and therefore in the amplitudes $\langle b|\psi_a(\Gamma)\rangle$) and on the amplitudes of transition between non degenerate states.

We compute the many-body mobility edge for the system in the following way. For each system size $N = 18, 20, 22, 24, 26$ we randomly generate a suitable number of realizations of disorder and for each of these we generate a set of initial states $a = a_1, a_2, \ldots, a_k$, making sure that their energy densities $\epsilon_a = E_a/N$ are (approximately) uniformly distributed in the range allowed by the model. Each of these states a is then propagated to its $\mathbb{Z}_2$-symmetric state $b = -a$ (global spin flip) using the forward approximation algorithm with fixed $\Gamma = 1$ in order to compute

$$\langle -a|\psi_a(\Gamma = 1)\rangle$$

Notice that the configuration $-a$ is the only configuration $b \in \{\pm 1\}^N$ to satisfy $\text{dist}(a, b) = N$, ergo for the state $|\psi_a\rangle$ we read from Eq. (4.7) that (after setting $r = N$)

$$\psi_N = \max\left\{|\langle b|\psi_a(\Gamma = 1)\rangle|\colon b \in \{\pm 1\}^N, \text{dist}(a, b) = N\right\} = |\langle -a|\psi_a(\Gamma = 1)\rangle|.$$

The results are then binned according to the energy density of the initial state a and the average of the random variable $Z_N = \ln|\psi_N|/N$ value was taken for each bin, obtaining $Z_N(\epsilon) = \langle Z_N\rangle$.
Using the formula $\Gamma_c(\epsilon) = \exp(-\ln|\psi_N(\epsilon)|/N) = \exp(-Z_N(\epsilon))$ from Eq. (4.9) we obtain a plot of the critical point for the localization transition as a function of the energy density ϵ shown in Fig. 4.5. Note that this is the energy density of the unperturbed eigenstates $|a\rangle$, while usually one would write Γ_c as a function of the energy density of the *perturbed* eigenstates $|\psi_a(\Gamma)\rangle$. However, the perturbed energies $E_a(\Gamma) = E_a + O(\Gamma^2)$ coincide with the unperturbed ones up to second-order corrections in Γ, which we neglect.

In order to plot the mobility edge in the (T, Γ)-phase diagram and compare it to the boundary of the glassy phase, we have to compute the relation $T = T(\epsilon)$ between temperature and (disorder-averaged) energy density. We used standard Monte Carlo methods to extract the (thermal) average energy density of different realizations of disorder at various temperatures (see Fig. 4.4) and fixed $\Gamma = 0$, then we took the average over the results[4]. In order to better understand the low energy regime we studied the ground state of the unperturbed (*i.e.* $\Gamma = 0$) model. For each size $N = 18, 20, 22, 24, 26$ we generated a large number (≥ 1000) of instances and extracted one of the ground states by performing a thermal annealing (whose results were checked against an exact solver for the smaller sizes). For each ground state we computed the Γ_c value

[4]We note that as one approaches the ground state energy, small difference of energies translate to (relatively) large difference of temperatures due to small values of the heat capacity in the low-temperature regime. In order to effectively control this effect one would require better precision in the M.C. energy estimation.

using the forward approximation. The disorder-averaged results are show in Fig. 4.6. Extrapolations give a value of $\Gamma_c = 0.67$ in the thermodynamic limit, which seems consistent with Fig. 4.5.
Finally, we plotted a finite-size ($N = 26$) estimate of the mobility edge and the line of the glassy transition in the (T, Γ)-phase diagram (Fig. 4.7). The localized phase seems to be strictly contained in the glassy phase, therefore there is a region of the phase diagram where the system is both glassy and delocalized.

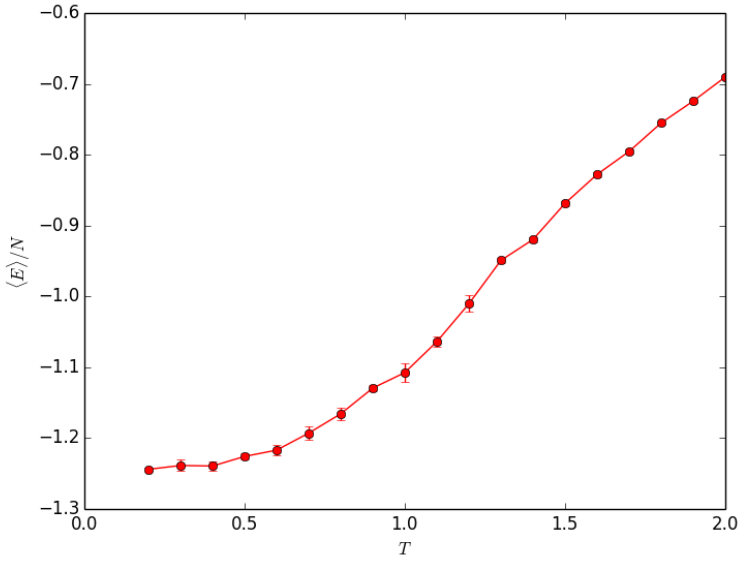


Figure 4.4: The disordered-averaged average energy density $\langle E \rangle / N$ at temperature T of the classical Ising spin glass Hamiltonian on the RRG of degree $d = 3$. The data was obtained using standard Monte Carlo techniques on systems of size $N = 150$.

4.5 Gibbs Averaging and Improved Analysis

One of the disadvantages of the method explained in the previous section is the necessity of mapping energies to temperatures in order to write the mobility edge $\Gamma_c(\epsilon)$ as a function of the temperature $\Gamma_c(T)$, so as to identify the localized phase in the equilibrium (Γ, T)-phase diagram of the model. This is a source of error in the data, as one can see from Fig. 4.4 that the very small value of the heat capacity $\partial \langle E \rangle / \partial T$ at low temperatures means that one needs increasingly better precision on the energy estimates in order to extract temperature differences with statistical confidence.

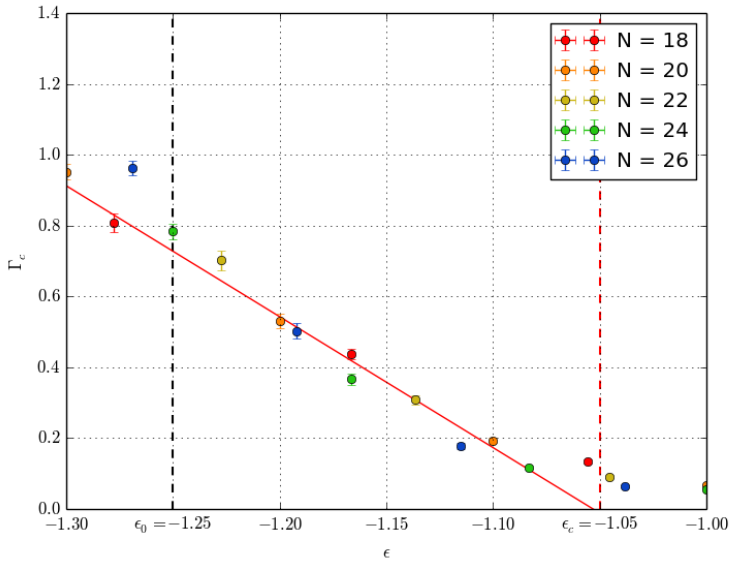


Figure 4.5: Finite-size estimates for Γ_c as a function of the unperturbed energy density ϵ, obtained using the forward approximation together with a linear fit of the data at largest N. The ground-state energy density of the unperturbed model in the thermodynamic limit is $\epsilon_0 \approx -1.25$, shown here as a dashed line while the critical energy density for the classical ($\Gamma = 0$) glassy transition is $\epsilon_c \approx -1.05$.

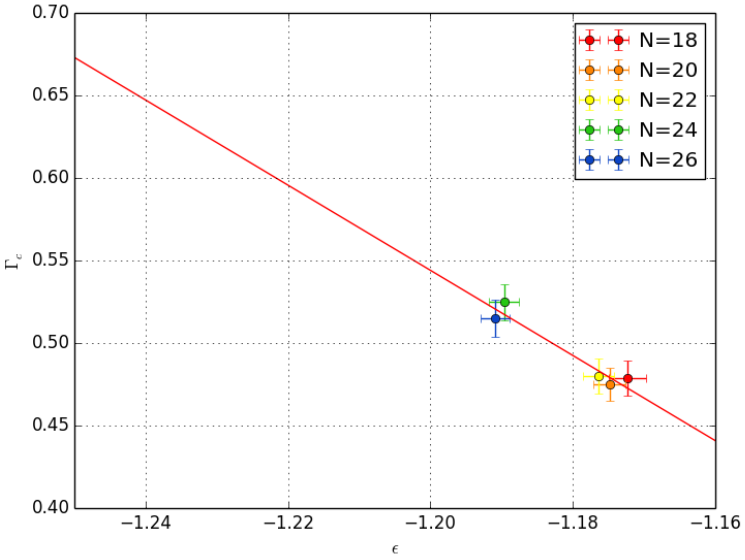


Figure 4.6: Estimates of the Γ_c for the ground state ($T = 0$ case) for different system sizes N. As $N \to \infty$ the disorder-averaged energy density of the ground state decreases towards the expected thermodynamic limit value of $\epsilon_0 = -1.25$. A linear interpolation of the Γ_c values obtained give a thermodynamic limit value of $\Gamma_c = 0.67$.

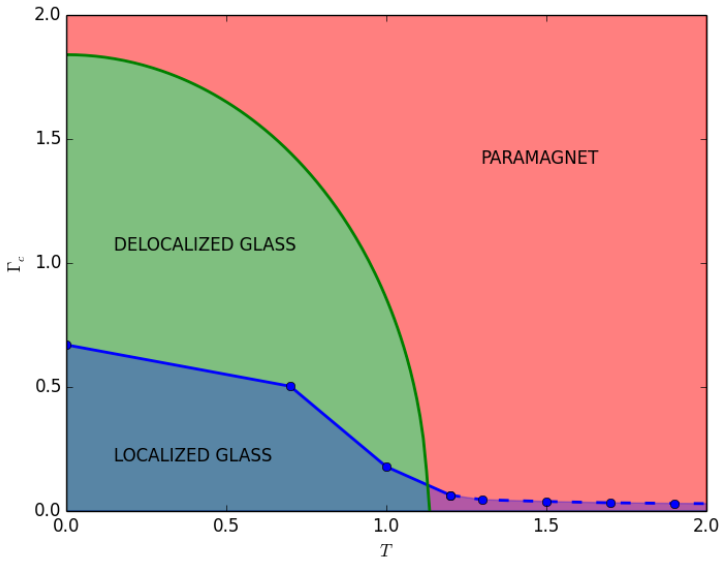


Figure 4.7: Phase diagram of the Hamiltonian of the Ising spin glass in a transverse field (Eq. (3.1)). The mobility edge obtained from the numerical data is shown as linked blue dots. The $T = 0$ point was obtained from the thermodynamic-limit extrapolation of Fig. 4.6 while the finite-temperature points were derived from the Γ_c values for the largest size ($N = 26$) of Fig. 4.5. With the possible exception of the $T > 1.1$ tail the localized phase seems to be a proper subset of the glassy phase.

One possible solution to this problem is to generate the initial states by sampling from the Gibbs distribution at temperature T and average the results, rather than generate some random states and then bin them according to their energy density ϵ. In this way one recovers imediately the critical value of Γ as a function of the temperature $\Gamma_c(T)$. So we proceed as follows. First we compute the thermal average of the value of Z_N over all the eigenstates $|E_n[J,\Gamma]\rangle$ of the Hamiltonian $H(J,\Gamma)$ of a given disorder realization J, for a fixed value of Γ:

$$\langle Z_N(J,T)\rangle \equiv \sum_n G\Big(E_n[J,\Gamma]\Big) Z_N\Big(|E_n[J,\Gamma]\rangle\Big), \tag{4.22}$$

where

$$G\Big(E_n[J,\Gamma]\Big) \propto \exp(-E_n[J,\Gamma]/T)$$

is the Gibbs probability given to the eigenstate $|E_n[J,\Gamma]\rangle$. Then we take its disorder-average

$$\overline{\langle Z_N(T)\rangle} = \sum_J p(J)\,\langle Z_N(J,T)\rangle\,. \tag{4.23}$$

Then using Eq. (4.9) with $\overline{\langle Z_N(T)\rangle}$ and $\Gamma = 1$ we can extract a finite-size estimator for the critical value $\Gamma_c(T)$ as a function of T. As before, we approximate the energies $E_n[J,\Gamma]$ by their unperturbed values $E_n[J,0]$ since these are corrected only at second order in Γ. Under this approximation, the computation can go as follows. For a fixed value of the system size N and temperature T we generate a number of disorder realizations $J_1,\ldots,J_m$. For each disorder realization J we use a classical Monte Carlo simulation to sample a number of states $|x_1\rangle,\ldots,|x_k\rangle$ which are typical for the given temperature T, *i.e.* they are sampled according to the Gibbs distribution at temperature T. Then using the forward scattering approximations we compute $\psi_N = |\langle x_i|\psi_{x_i}\rangle|$ and through averaging of the quantity $Z_N = \log(\psi_N)/N$ over the results we obtain $\langle Z_N(J,T)\rangle$ of Eq. 4.22. Then we take the average over the disorder and obtain $\overline{\langle Z_N(T)\rangle}$ and $\Gamma_c(T) = \exp(-\overline{\langle Z_N(T)\rangle})$. The results are shown in Fig. 4.9. Figures 4.10,4.11 and 4.12 show how the distribution of the random variable $Z_N(T)$ changes as one increases the size of the system, for fixed temperatures $T = 1.1, 0.7, 0.5$. These three temperatures examplify three behaviours of the $\Gamma_c(N)$ curves seen in Fig. 4.9: at $T = 1.1$ the $\Gamma_c(N)$ curve is decreasing with N, at $T = 0.5$ the curve is increasing with N, and at $T = 0.7$ the curve increases, then reaches a maximum and then decreases.

Localization transitions are dynamical rather than thermodynamical transitions, so we cannot use the finite-size scaling Ansatz to study this transition. However, we can always try to find a universal curve that fits the data. In order to do this, the first step is to obtain a data collapse of the curves at different temperatures by some temperature-dependent transformation φ_T applied to the data points $(1/N, \Gamma_c(N,T))$ that make up the $\Gamma_c(N)$-curve for temperature T. This was accomplished in essentially the following way. By starting with the curve of $T = 0.9$ as reference, we tried to rescale the coordinates of the next curve, $T = 1.1$ so as to make it overlap with the first. We discovered that by independently rescaling the two coordinates $\Gamma_c(N,T)$ and $1/N$ by factors $A \approx 1.8$ and $B \approx 1.125$

$$\phi\Big(\Gamma_c(N,T), \frac{1}{N}\Big) = \Big(\frac{\Gamma_c(N,T)}{A}, \frac{B}{N}\Big),$$

one could obtain the desired collapse of the two curves. We repeated the procedure for the other curves and obtained a set of temperature-dependent scaling factors $A(T)$ and $B(T)$ that produce a full data collapse (see Fig. 4.13) of the curves in Fig. 4.9. The transformation φ_T for a generic T is then given by

$$\varphi_T\Big(\Gamma_c(N,T), \frac{1}{N}\Big) = \Big(\frac{\Gamma_c(N,T)}{A(T)}, \frac{B(T)}{N}\Big). \tag{4.24}$$

The fact that the values of $A(T), B(T)$ are well fitted by simple, monotonic functions (shown in Fig. 4.8) is a sanity check that our method is not arbitrary.

Once the data collapse was obtained we tried to find a curve through the collapsed data. We used a modified Padé approximant [79] (*i.e.* a ratio of two polynomial functions of given orders) with the coefficients of the polynomials as fitting parameters. We discovered that one can fit the collapsed data with a universal curve $f(x)$ given by

$$f(x) \equiv x\frac{1}{1+Cx^{3/2}} \tag{4.25}$$

where $C = 4.91$ is a fixed parameter. Fig. 4.14 shows the data collapse and the universal curve. While we stress the fact that this is just a tentative interpretation of the data, and should not be taken as definitive, this data collapse suggests the possibility that the MBL region disappears in the thermodynamic limit, collapsing onto the $\Gamma = 0$ line of the phase diagram.

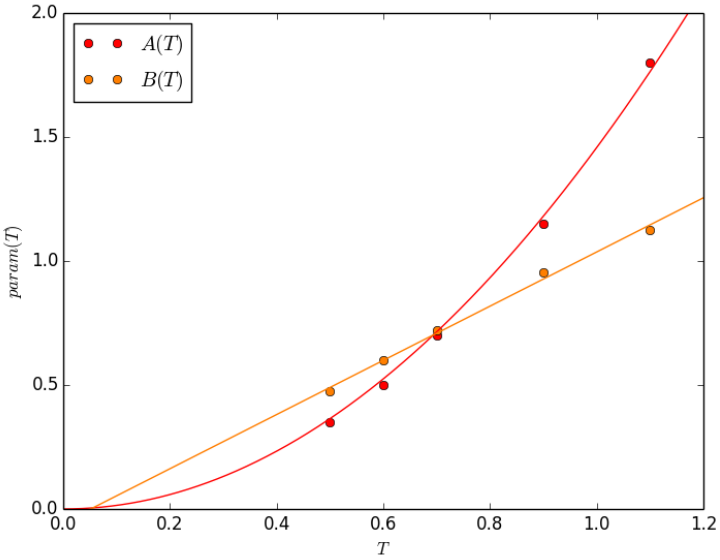


Figure 4.8: plot of the temperature-dependent dilation parameters $A(T), B(T)$ for the rescaling of (respectively) the Γ_c and the $1/N$ coordinates according to Eq. (4.24)

.

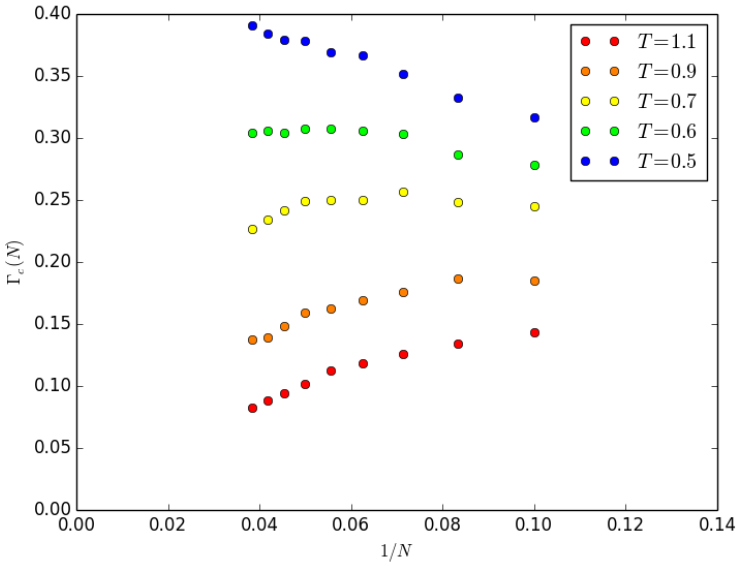


Figure 4.9: finite-size estimates for the critical point $\Gamma_c(N)$ of the localization transition, for different sizes $N = 10, 12, 14, \ldots, 26$ and temperatures T.

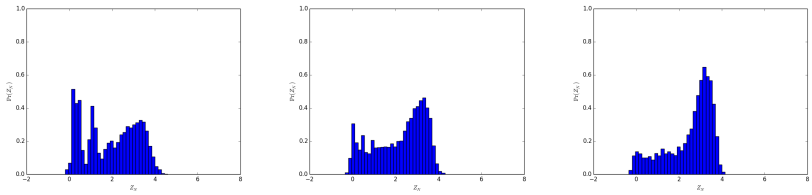


Figure 4.10: Distribution of the random variable $Z_N(T)$ for fixed temperature $T = 1.1$ and increasing system sizes $N = 10, 18, 26$ (increasing from left to right). Note the formation of an asymmetric peak with a comparatively larger left tail. This behaviour is typical of temperatures whose $\Gamma_c(N)$ curves decrease monotonically with N within a given range of system sizes (see Fig. 4.9).

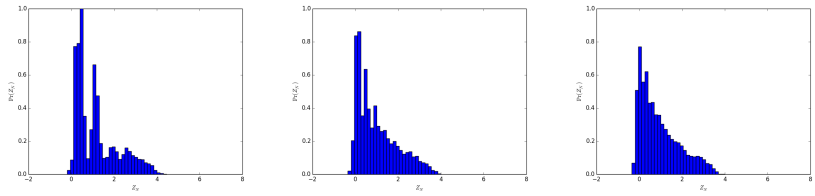


Figure 4.11: Distribution of the random variable $Z_N(T)$ for fixed temperature $T = 0.5$ and increasing system sizes $N = 10, 18, 26$ (increasing from left to right). Here as well we see the appearance of an asymmetric peak. In this case however the larger tail of the distribution is the right one. This behaviour is typical of temperatures whose $\Gamma_c(N)$ curves increase monotonically with N within a given range of system sizes (see Fig. 4.9).

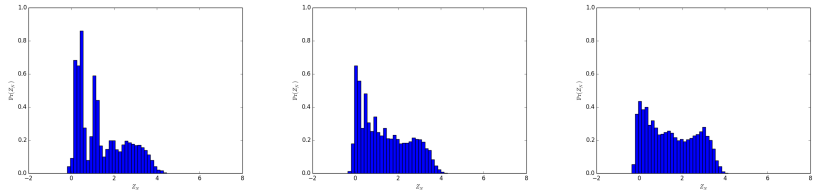


Figure 4.12: Distribution of the random variable $Z_N(T)$ for fixed temperature $T = 0.7$ and increasing system sizes $N = 10, 18, 26$ (increasing from left to right). The distribution starts similar to the $T = 0.5$ case but develops a second peak at larger sizes. This behaviour is consistent with the data of Fig. 4.9, where the curve for $T = 0.7$ is seen to grow at small system sizes, reach a maximum and then drop at larger system sizes. Analogously, the distribution of $Z_N(T)$ seems to be connecting the behaviour of the $T = 0.5$ case to the behaviour of the $T = 1.1$ case.

4.6 Many-body Localization and Level Spacing

Anderson localization is the theory of *non-interacting* electrons in random potentials. This looks like a very drastic oversimplification, as electrons cannot be shielded from the Coulomb interaction. The analogue of Anderson localization where electron-electron interactions are not neglected goes under the name of *many-body localization* (MBL). Originally discussed in the early days of Anderson localization, it has recently received great attention after Ref. [80] shown that localization is not destroyed by weak interactions. MBL theory was further expanded to include more generic systems than just electrons hopping on a lattice and was also extended to apply also for disordered spin systems [81, 82, 83].

While the representation of the transverse-field Ising spin glass in the Eq. (4.2) bears a close resemblance to the Anderson model – and therefore suggests Anderson localization – the physical system is made of interactiong spins, so it is probably more proper to call the localization seen in the previous section an example of many-body localization. What matters for the Adiabatic Quantum Algorithm are the spectral and dynamical properties of the Hamiltonian $H(\Gamma)$,

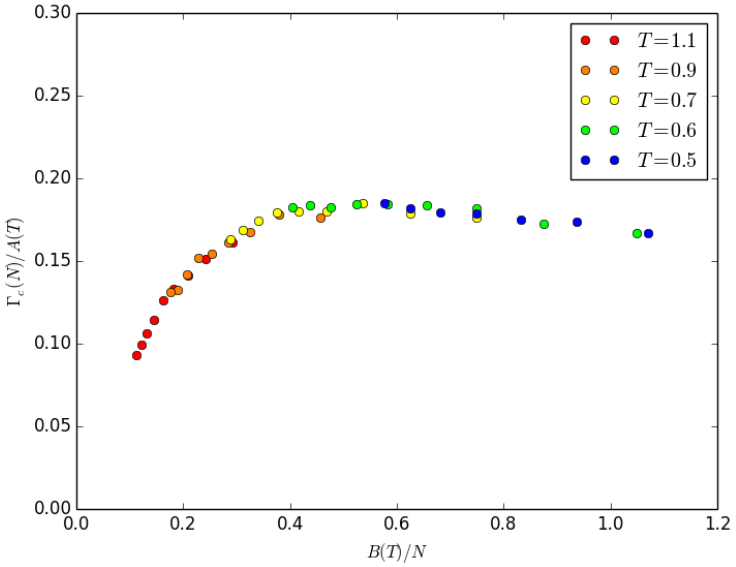


Figure 4.13: Data collapse for the Ansatz given by Eq. (4.24).

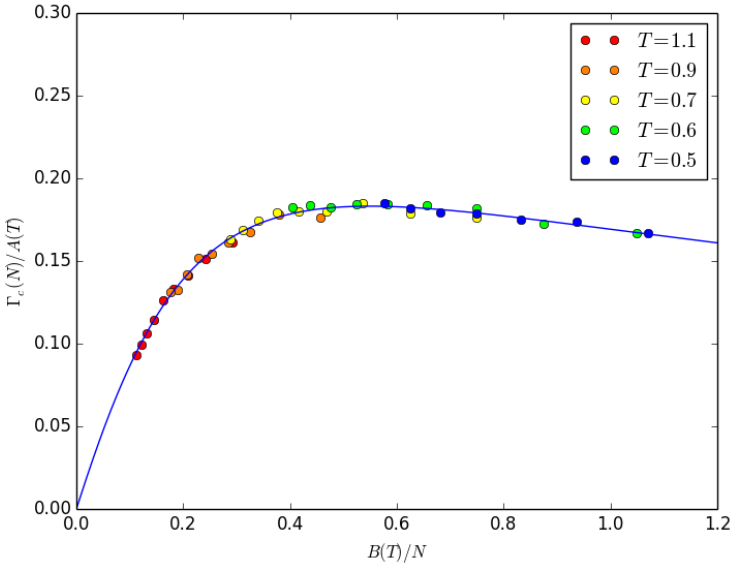


Figure 4.14: Universal curve defined by Eq. (4.25) shown on top of the data collapse of Fig. 4.13. If correct, this interpretation would show that the MBL exists only at finite system sizes, at least according to the forward scattering approximation. It is possible that the self-energies neglected in the FSA could shift the Γ_c values and restore localization.

and these are independent of any choice of basis.

Many-body localization is the robust breaking of ergodicity due to the emergence of quasi-local integrals of motion which constrain the dynamics in a way similar to how conserved local charges do in the case on integrable systems. Quantum ergodicity is often studied by looking at the spectral statistics of the model's Hamiltonian. The justification for this approach is to be found in the connection between classical systems and their quantized counterparts. There are two main results to consider: firstly, we have the Bohigas-Giannoni-Schmit (BGS) conjecture [84] which states that the spectra of (time-reversal invariant) quantum systems whose classical analogues are ergodic exhibit the same statistics of the Wigner-Dyson ensembles. This entails in particular that the joint probability distribution for the energy eigenvalues of such quantum Hamiltonians is of the form

$$P(E_0, \dots, E_{N-1}) = \mathcal{N}_1 \prod_{i<j} |E_i - E_j|^\beta \exp(-\mathcal{N}_2 \sum_i E_i^2),$$

where $\mathcal{N}_1, \mathcal{N}_2$ are normalization constants and the parameter $\beta \geq 1$ depends on the specific random matrix ensemble which correctly describes the symmetries of the Hamiltonian[5]. The term $|E_i - E_j|^\beta$ is indicative of a phenomenon typical of ergodic systems called *level repulsion*, as spectra with energy levels very close together, $|E_i - E_j| \ll 1$, receive a small probability.

There is an analogous conjecture that applies to integrable systems. The Berry-Tabor (BT) conjecture [85] states that the level spacing in the spectra of quantum systems whose classical analogues are integrable exhibit a Poisson-like distribution, and level repulsion is absent.

While these are indeed conjectures, there is a great deal of numerical evidence supporting them, even though a few exceptions are known [86, 87]. It is common to take them to be the *definition* of ergodicity (or ergodicity breaking) of the states belonging to the given energy shell in quantum system [88]. The naive approach to decide whether a system is ergodic or ergodicity-breaking would be to generate a large enough number of disorder realizations, extract the full spectrum using *e.g.* the Lanczos method, and try and fit the level spacing statistics with either a Wigner-Dyson or a Poisson distribution.

However, since the system we want to study has a mobility edge, the expectation is that ergodicity is mantained at all the energies corresponding to temperatures outside of the localized region in Fig. 4.7, while inside of the localized region one should witness the breaking of ergodicity. For this reason we need a way to check ergodicity inside of a given energy interval. A way to do this was proposed in [89], where the authors proposed the average *spacing ratio* $\overline{r}$ as a witness for ergodicity. If $E_0, E_1, \dots$ are the energies of the system, then one defines the sequence of spacing ratios $\{r_n\}$

$$r_n \equiv \frac{\min(\Delta_n, \Delta_{n+1})}{\max(\Delta_n, \Delta_{n+1})}$$

where $\Delta_n \equiv E_n - E_{n-1}$. This quantity is then computed for the energy levels in a small energy shell and averaged with respect to the realizations of disorder,

[5] the most common ensembles are the Gaussian Orthogonal Ensemble (GOE) with $\beta = 1$, the Gaussian Unitary Ensemble (GOE) with $\beta = 2$, and Gaussian Symplectic Ensemble (GOE) with $\beta = 4$.

obtaining $\overline{r}(E)$. It is known that $\overline{r} \approx 0.39$ for a Poisson distribution while $\overline{r} \approx 0.53$ for a Wigner-Dyson distribution [89]. One would then use this fact to check whether, for a given energy E, the system is ergodic in the energy shell $(E, E + \mathrm{d}E)$ by averaging the spacing ratio r_n for energy levels inside of this interval.

When applied to the transverse-field Ising spin glass, this technique proved unsatisfactory for a number of reasons. The main bottleneck of the method is the necessity to extract the full spectrum of the Hamiltonian, as this strongly limits the largest system size one can access. Note that at $\Gamma = 0$ the spectrum of the Hamiltonian (3.1) is given by the classical (σ^z-diagonal) spin glass term. The disordered interactions $J_{ij} = \pm J$ being of the Ising type, these energy levels are usually highly degenerate, because by going back to the COP picture one can see that these count the number of constraints violated by a given classical spin configuration. The spectrum $\mathrm{Sp}(H)$ is then composed of integer multiples of J. It is easy to see that on a regular graph of degree $d = 3$, the energy densities populated by energy levels are contained in the set

$$\Big[-J\frac{3}{2}, J\frac{3}{2}\Big] \cap J\mathbb{Z}.$$

For a system of size N the separation between two adiacent values of the energy density is $3J/N$, so that for two consecutive sizes N and $N+2$ the two sequences of possible values for the energy density $\{\epsilon_i^{(N)}\}$ and $\{\epsilon_i^{(N+2)}\}$ are shifted, and have points that lie in between one another, so that extracting a large-N limit of any function of the energy density is hard unless one goes to very large system sizes where the available energy density values accumulate inside of a given window of energy density.

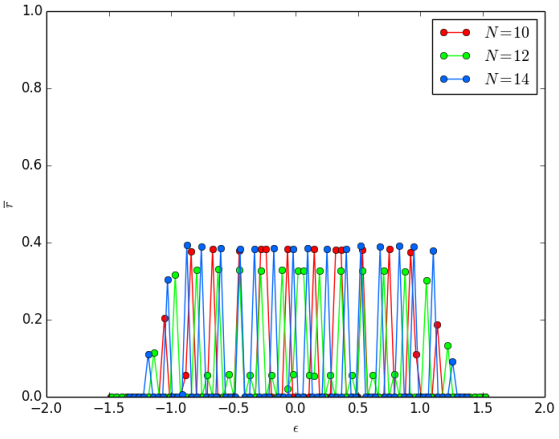
To each energy level of the unperturbed $\Gamma = 0$ Hamiltonian is associated a cluster of classical spin configurations, all with the same energy. When the transverse field term is turned on and $\Gamma > 0$, these clusters split into a number of different energy levels but, even for $\Gamma \approx J$ these cluster are still fairly well-separated. Therefore the same issues with extracting a thermodynamic-limit behaviour from the numerical data remain.

Finally, the temperatures at which the system is in the glassy phase correspond to a very small energy interval above the ground state energy, as one can see for example from the "temperature-to-energy density" conversion table we have computed for large ($N = 150$) at $\Gamma = 0$ (see Fig. 4.15), so that the energy density interval associated to glassy dynamics lies at the edge of the spectrum that is pathologically affected by large fluctuations of the value $\overline{r}$ (see Fig. 4.15(b)).

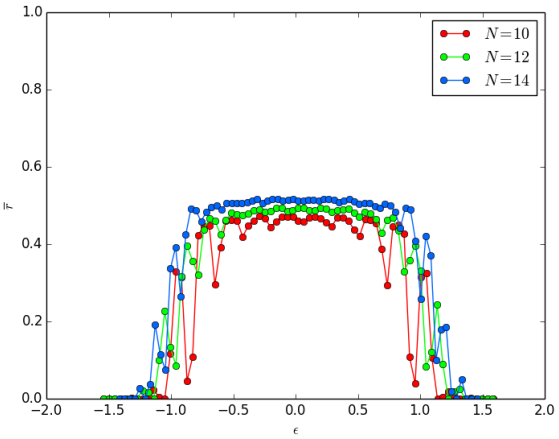
For the sake of completeness, let us try anyway with two values of Γ, one that lies completely outside of the glassy phase, say $\Gamma = 2.0$ and another where the system is expected to have temperature-driven localized/delocalized transition. In the first case we expect the system to be ergodic at essentially any temperature T (see the phase diagram of Fig. 4.7), while in the second case this will be true only at high temperatures. As a first sanity check, note that for both values of Γ, if we look at the behaviour of the average spacing ratio around the $\epsilon = 0$ energy density (which is the energy region associated to high temperatures) the value is monotonically increasing (see Fig. 4.16) and is at least compatible with a possible convergence to the expected value of $\overline{r} \approx 0.53$.

However, we were unable to reproduce even this very weak result when we tried to probe the low energies associated to localization. For example, according to our estimate given in Fig. 4.5, if $\Gamma = 0.6$ then the energies densities at which the system is localized should lie approximately between $\epsilon = -1.2$ and $\epsilon = -1.25$. Within this narrow interval the statistics we were able to accumulate (whose $\overline{r}$ values are shown in Fig. 4.17) is extremely scarse consisting in < 10 samples per point, and even though one notices a slight monotonic behaviour with N, the differences between distinct data points are so small and their values so remote from the expected value of $\overline{r} \approx 0.39$ that the information we can extract about their convergence is essentially nil.

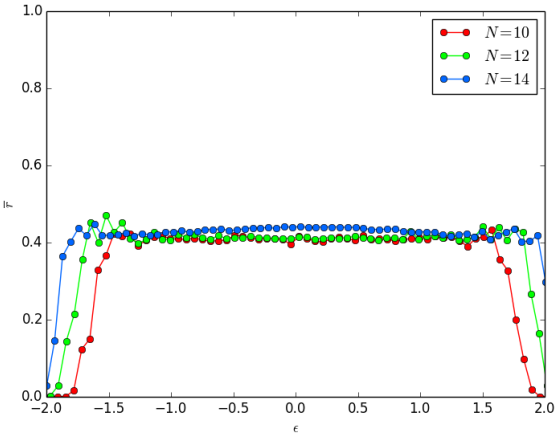
For these reasons, the final results is that the spacing ratio analysis we attempted is unfortunately inconclusive isofar as establishing ergodicity, or lack thereof, is concerned.



(a) Spacing ratios at $\Gamma = 0$ (classical spin glass). Note the well-separated peaks associated to the degenerate, integer-valued energy levels.



(b) Spacing ratios at $\Gamma = 0.6$ (inside of the glassy phase). The clusters spread and start fusing together. However there are strong fluctuations at the boundary of the spectrum.



(c) $\Gamma = 2.0$ (ergodic phase)

Figure 4.15: Spacing ratios of the spectrum of the Hamiltonian (3.1) at different values of the transverse-field strength Γ.

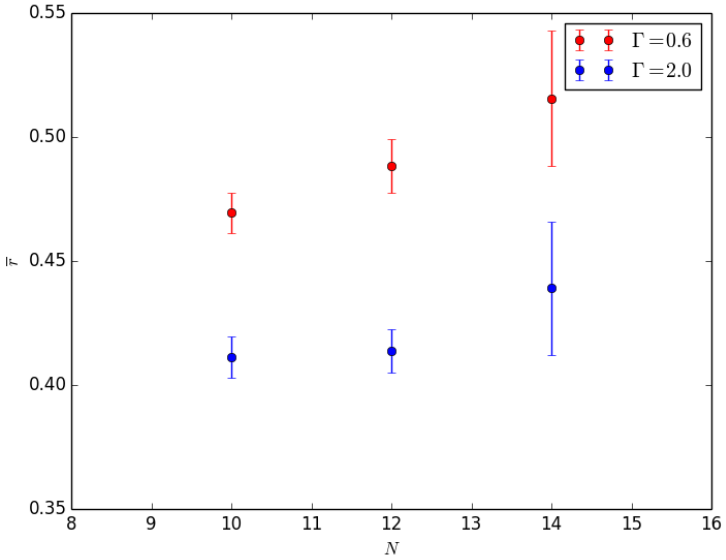


Figure 4.16: The average spacing ratios of the Hamiltonian $H(\Gamma)$ for two values of $\Gamma = 0.6, 2.0$ at a fixed energy density $\epsilon = 0.1$ and system sizes $N = 10, 12, 14$.

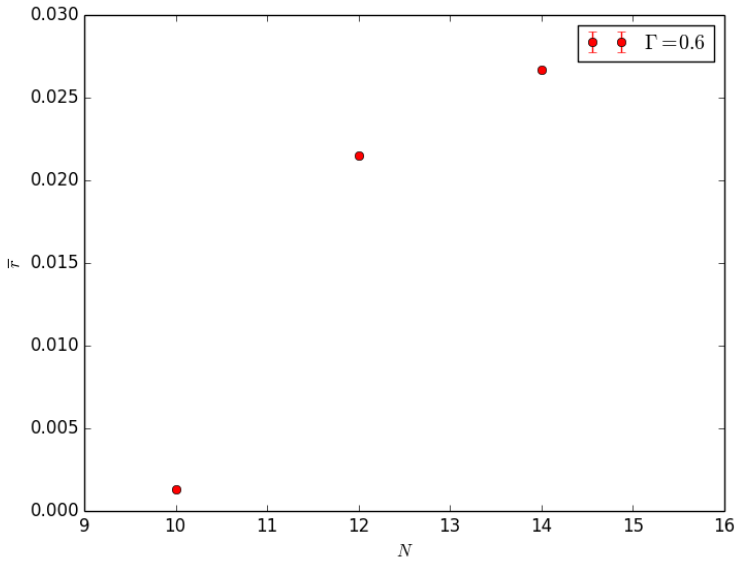


Figure 4.17: The average spacing ratios of the Hamiltonian $H(\Gamma)$ for $\Gamma = 0.6$ inside of the energy density interval $[-1.25, -1.2]$ for system sizes $N = 10, 12, 14$. Even though the error bars are smaller than the data points, they are likely to be unreliable due to the very scarce statistics available.

Chapter 5

Conclusions

The goal of this thesis was to study the Adiabatic Quantum Algorithm from the perspective of statistical physics and condensed matter physics. In order to accomplish this task we chose to focus on the transverse-field Ising spin glass model defined on a regular random graph of degree $d = 3$. This classical Ising spin glass model is widely believed to capture the complicated combinatorial structure of general NP-hard computational problems. As the strength Γ of the transverse field is taken from infinity to zero, this defines a common choice for the adiabatic path of the Adiabatic Quantum Algorithm called quantum annealing. From the point of view of quantum statistical mechanics, this path represents the $T = 0$ line in the (T, Γ)-plane of the model. Before our work, this line was the final piece that was missing from its phase diagram, which was pieced together in a number of previous works. With this contribution the picture of the phase diagram is now complete. We computed numerically the disorder-averaged Rényi entropy $S_A^{(2)}$ when the region A is taken to be half of the system. We focused on the paramagnetic phase up to the critical point. We found that the Rényi entanglement entropy satisfies a volume law for all values of the transverse field Γ we considered, with a prefactor that is maximal at a point we interpret as the critical point of a QPT. We saw that the critical point Γ_c we extracted from the Rényi entropy data coincides within statistical errors with the critical point of the glassy phase transition of the model, which we found by studying the Edwards–Anderson parameter q_{EA}. We also saw that q_{EA} is continuous at this critical point.

We studied the disorder-averaged values of two (spatial) two-point correlation functions of the model — the maximal correlation and the mean correlation — in order to get a better understanding of this phase transition. The decay of both of these correlation functions was found to be compatible with a stretched-exponential for all values of Γ, both critical and off-critical. We extracted critical correlation lengths $\xi_N(\Gamma_c)$ and observed that they decrease with the system size N and converge to a finite value at the critical point of the glassy transition. We conclude that this phase transition exhibits features of both first-order transitions (finite critical correlation length) and second-order ones (continuous order parameter). From the four above-mentioned quantities we extracted the following estimate for the critical point: $1.82 \leq \Gamma_c \leq 1.85$.

Small-size numerical studies of the quantum Fisher information point to the fact that critical and off-critical multipartite entanglement are microscopic, *i.e.*

they never involve more than a number of spins that is constant in the system size. This might *prima facie* seem to contrast with the volume-law of the Rényi entropy but we believe that the following simple picture of the wavefunction to be compatible with all the data. At $\Gamma = \infty$ the wavefunction is a product state of σ^x eigenstates. As Γ is decreased, pairs of nearest-neighbours entangled spins are created and become increasingly entangled as Γ approaches the critical point of the QPT. The volume-law scaling of the Rényi entropy is a consequence of the expander-graph structure of the interaction graph of the model: any bipartition A cuts through an extensive number of these 2-spin entangled states, each of which gives a finite contribution to the entanglement entropy $S_A^{(2)}$.

In order to further explain these results we attempted a perturbative calculation using the large-Γ transverse-field term as the unperturbed Hamiltonian and the spin-glass coupling strength J as the perturbative parameter. The spectrum of the unperturbed Hamiltonian can be interpreted as a vacuum of quasiparticles, on top of which there are equally-spaced bands of increasing quasiparticle density. First-order parturbation theory in J showed that the energy of the first excited state (belonging to the one-particle band) crosses the ground state energy at $\Gamma_c \approx 1.414$. Going to higher bands and using a mean-field Hamiltonian that reproduces some of the contributions of higher-order perturbation terms we see that the critical points moves to higher values of Γ, in better agreement with the numerics. The idea is then that at the critical point one of the higher bands crosses the vacuum energy of the quasiparticles, so that energy considerations favour the creation of a finite density of quasiparticles. This perturbative approach is also corroborated by the fact that its predictions are consistent with the known results for the limit $d \to \infty$ of infinite degree, where the RRG Ising spin glass goes to the Sherrington-Kirkpatrick model.

We also studied the localization properties of the model in the limit where the transverse-field is weak compared to the disordered interactions. This regime describes the final stage of a quantum annealing protocol designed to find the ground-state energy of the Ising spin glass. Many-body localization has been argued to be an obstacle to efficient quantum annealing due to the presence of exponentially-closing gaps in the localized phase.

We computed numerically the many-body mobility edge of the system in the forward approximation, finding that the energy eigenstates of the system indeed localize for small values of the transverse field at finite system sizes. When plotted against the equilibrium phase diagram of the model, we discovered that the localized region does not coincide with the glassy phase. In particular, evidence points to the fact that the glassy phase is partitioned into a delocalized region and a localized one.

Let us examine these results with respect to the background literature. From the point of view of quantum computation we note that the microscopicity of the entanglement, along with the fact that exponentially-closing avoided crossings in the paramagnetic region are highly unlikely, suggest on the one hand that quantum annealers do not need to generate and sustain an inordinate amount of entanglement in order to correctly follow the adiabatic path, at least up to the critical point of the transition. On the other hand, this part of the adiabatic-evolution dynamics can be efficiently simulated on a classical computer using the standard methods of Refs. [90, 40]. The extension of these considerations to the entire adiabatic path rests upon the scaling behaviour of the multipartite

entanglement and the minimal gap inside of the glassy phase. Previous results on the related problem 3-regular MAX-CUT (which is equivalent to our model without disordered interactions J_{ij}) found that the gap closes superpolynomially fast inside of the glassy phase. We expect this to be the same also in the Ising spin glass. Also, classical considerations suggest that at least for $\Gamma = 0$ we should have ground states that are superpositions of (quantum states that represent) the solutions to the classical problem, *i.e.* the classical Ising spin glass. These quantum states are expected to be products of smaller entangled and unentangled states. For example, if "000101" and "000011" are the only two solutions to the classical problem, then their superposition can be written as

$$|000101\rangle + |000011\rangle = |000\rangle \otimes (|10\rangle + |01\rangle) \otimes |1\rangle$$

that is, a six-qubit state that is only 2-entangled. Therefore entanglement is dependent on the structure of the solutions to the classical problem, which was studied extensively [91, 92]. Nevertheless, we are unable to derive the disorder-averaged entanglement properties of the quantum ground states given the current state of knowledge.

The presence of a MBL region inside of the glassy phase model is consistent with the predictions of the Altshuler *et al.* paper [67]. However, while our numerical data show the existence of a localized phase at finite system sizes, the data collapse we presented at the end of Chapter 4 suggests this region shrinks more and more as one approaches the thermodynamic limit. While finite-size properties are the ones that are really relevant for applications in real-life quantum computers, one has to admit that the extension and indeed the existence of the MBL phase in the thermodynamic limit is still open to question as we are currently lacking a proper finite-size scaling of the mobility edge.

We believe there are several future directions that are worth exploring. For the transverse-field Ising spin glass model, an obvious one is to extend our study of the Rényi entropy inside of the glassy phase, where ergodicity breaking manifests itself as a fast increase of the convergence time needed by the Monte Carlo simulations. Methods such as parallel tempering have been used to ameliorate this effect so it is reasonable to believe that this obstacle is not an insurmountable one. Further, a more systematic study of the mean-field approach and of the developement of quasiparticle descriptions of the physics of adiabatic paths seem useful in order to develop a better intuition.

For the MBL part, the first *desideratum* would be to try and confirm the predictions of the forward scattering approximation. Self-energy corrections to the FSA can be computed using a similar, albeit more resource-heavy version of the transfer matrix method we described and used in Chapter 4. Another natural future direction would be to check whether the same localization/delocalization transition is present when the disordered term of the Hamiltonian encodes a real-life computational problems such as 3-SAT. In the affirmative case, a detailed comparison of the performance of *e.g.* simulated annealing and quantum annealing (either simulated numerically or by an actual experiment) inside of the region that is both glassy and delocalized would help shed light on the realistic capabilities of quantum annealers over classical thermal annealing and other algorithms based on stochastic local optimization. Moreover, while our work considered the eigenstate localization the dynamical properties of the glassy,

delocalized phase are at this point unknown. This is worthy of further investigation as we expect that classical methods that exploit the fine-tuning of thermal relaxation (such as simulated annealing) will perform poorly in the entire glassy phase while quantum annealers will perform poorly only once (dynamical) localization sets in. In case the glassy, delocalized region of the phase diagram should turn out to be ergodic (*e.g.* thanks to quantum tunnelling) then this could constitute an example of a region of the phase diagram where quantum annealing algorithms can be expected to outperform any classical thermal annealing protocol. This would be solve an open problem that has troubled the quantum community in the past ten years.

For the larger picture of Adiabatic Quantum Computing, a better understanding of the entanglement in the ground-state wavefunctions of the Quantum Adiabatic Algorithm seems auspicable. In particular, when does an adiabatic path encounters points of macroscopic entanglement (*i.e.* entangled blocks of size $O(N)$, as measured by the quantum Fisher information), and how does this feature relate with the minimal gap? From a computational perspective, there are many reason to believe that quantum computing is strictly more powerful than classical computing, so at least is some cases the Adiabatic Quantum Algorithm should be impossible to simulate efficiently on a classical computer. Currently, the only known way for this to happen is if multipartite entanglement should scale extensively at least at some point of the adiabatic path. Quantum algorithms that efficiently solve problems that do not have efficient classical algorithms should show this behaviour. On the other hand, if a quantum algorithm is already inefficient (*e.g.* it has a gap that closes exponentially fast), then no conclusions on the entanglement follow. One could then ask what happens in this case. Is extensive multipartite entanglement simply a typical feature of all (or most) adiabatic runs, or does it correlate with the minimal gap? It seems that further research is needed to address this point.

In closing, we hope that this thesis has shown adiabatic quantum computing to be a useful framework for the analysis of quantum algorithms, where physical intuitions, concepts and techniques can be brought to bear profitably.

Appendix A

Appendices

A.1 Perturbation Theory by Transfer Matrix

As mentioned in Chapter 2, the adjacency matrix A_G of an undirected graph $G = (V, E)$ is defined by

$$\langle i|A_G|j\rangle = \begin{cases} 1 & \text{if } (i,j) \text{ is an edge in } E \\ 0 & \text{otherwise} \end{cases}.$$

It is a well-known result in algebraic graph theory that the powers of the adjacency matrix A_G of an undirected graph G contain information about the paths in the graph, specifically

$$\langle i|(A_G)^n|j\rangle = \text{number of paths of length } n \text{ connecting the vertices } i \text{ and } j$$

This can be generalized for directed, weighted graphs: in this case each directed edge $e = (i \to j)$ of G will be associated to a weight $w(e)$. The weight of a path $p = (e_1, \ldots, e_k)$ (a sequence of pairwise incident edges of G) is defined as the product of the weights of its edges:

$$W(p) \equiv \prod_{i=1}^{k} w(e_i)$$

In this case we can define the *transfer matrix* T

$$\langle i|T|j\rangle = \begin{cases} w(e) & \text{if } e = (j \to i) \text{ is a directed edge in } E \\ 0 & \text{otherwise} \end{cases}.$$

Now the powers of T encode sum-over-paths quantities in G:

$$\langle i|(T_G)^n|j\rangle = \sum_{p \in \text{paths}(j,i;n)} W(p) = \sum_{p \in \text{paths}(j,i;n)} \prod_{e_i \in p} w(e_i) \tag{A.1}$$

where the sum is taken over all paths p in G of length n (*i.e.* composed of exactly n edges) connecting j to i. One can use transfer matrices over weighted graphs to compute the forward-scattering approximation (FSA) of Eq. (A.2).

In Section 4.3.1 we needed to compute the amplitudes of the perturbed energy eigenstates $|\psi(\Gamma)\rangle$ of a transverse-field Hamiltonian $H(\Gamma)$

$$H(\Gamma) = H_0 - \Gamma \sum_i \sigma_i^x$$

where H_0 is diagonal in the σ^z basis. For eigenstates $|x\rangle, |y\rangle$ of H_0, using the FSA we have

$$\langle y|\psi_x(\Gamma)\rangle \approx \sum_{p\in \text{paths}(x,y,n)} \prod_{i\in p} \frac{\Gamma}{\epsilon_x - \epsilon_i}, \tag{A.2}$$

where $n \equiv \text{dist}(x, y)$ is the Hamming distance between x and y (the number of spin flips that connectes them). Again, we indicate with $|\psi_x(\Gamma)\rangle$ the eigenstate of $H(\Gamma)$ that is adiabatically connected to $|x\rangle$ in the limit $\Gamma \to 0$.

The r.h.s of Eq. (A.2) can be seen as a special case of Eq. (A.1) where the underlying graph G is the N-th dimensional Boolean hypercube $\mathbb{B}^N = \{0, 1\}^N$ and the weight of an edge $e = (j \to i)$ of the graph is given by

$$w(j \to i) = \frac{\Gamma}{\epsilon_x - \epsilon_i}.$$

We see then that the transfer matrix $T \equiv T_x$ is

$$T_x = D_x A,$$

where A is the adjacency matrix of the Boolean hypercube and D is a matrix defined by $\langle i|D_x|j\rangle = \delta_{ij}\Gamma/(\epsilon_x - \epsilon_i)$. Then Eq. (A.2) is equal to $\langle y|T_x^d|x\rangle$ where $d = \text{dist}(x, y)$. Therefore we can compute the amplitude $\langle y|\psi_x(\Gamma)\rangle$ by starting in the initial state $|x\rangle \equiv |x_0\rangle$ and repeatedly applying the transfer matrix, defining vectors $|x_0\rangle, |x_1\rangle, \ldots$

$$|x_{t+1}\rangle \equiv T|x_t\rangle \qquad \text{for } t = 1, 2, \ldots, n.$$

Then $\langle y|\psi_x(\Gamma)\rangle = \langle y|x_n\rangle$.

Efficient implementation

Of course $|x_t\rangle$ is a vector with 2^N entries and T_x is a $2^N \times 2^N$ matrix with real coefficients, so their sizes are exponential in N. However, one can use a number of tricks to implement the procedure more efficiently.

1. **Reduced transfer matrix.** When written in the σ^z basis, the vector $|x_t\rangle$ is very sparse during most of the computation, so at each step most entries of the transfer matrix T are irrelevant (since they will be multiplied by the zero entries of $|x_t\rangle$). This is because the transfer matrix T connects only neighbouring states on the hypercube, so if we start in a basis state $|x_0\rangle \equiv |x\rangle$ (totally localized), then $T^k|x\rangle$ will be supported on states inside of a ball of radius k centered in x:

   $$T^k|x\rangle = \sum_{z:\text{dist}(x,z)\le k} \alpha_z|z\rangle.$$

   Moreover, since we need to propagate only for $n = \text{dist}(x, y)$ steps, the final state y at the final time $t = n$ will be exactly on the boundary of

such a ball. This means that at each step we need only to propagate the *boundary* of this ball: the amplitude of the states in its interior are irrelevant for the FSA. This means that

- the vector $|x_t\rangle$ requires only $\binom{N}{t}$ real-valued entries to be described completely, instead of 2^N.
- one does not need to store in memory the entire transfer matrix T_x. At each step t one needs only a reduced transfer matrix $T_{x,t}$ to propagate $|x_t\rangle$ to $|x_{t+1}\rangle$. This matrix has only $(N-t)\binom{N}{t}$ non-zero entries, instead of $N2^N$ of T_x.

2. **Propagation cost (memory).** At the t-th propagation step we have to compute the vector $|x_{t+1}\rangle$ from the vector $|x_t\rangle$ by applying the (reduced, directed) transfer matrix $T_{x,t}$. Both vectors have to be stored in memory during a single propagation step and only when we have completed computing $|x_{t+1}\rangle$ can then $|x_t\rangle$ be discarded. Combined together, the two vectors have $\binom{N}{t} + \binom{N}{t+1} = \binom{N+1}{t+1}$ non-zero entries. The transfer matrix $T_{x,t}$ can be computed on the fly entry by entry and does not use any appreciable amount of memory. The bottleneck is then around the middle of the propagation, where $t+1 \approx (N+1)/2 \in \exp(N)$. This is confirmed by the data from the simulations we did using the ULYSSES cluster, where an exponential fit gives a memory requirement (for the full propagation on the entire hypercube of dimension N) of

$$\mathrm{Mem}(N) = \alpha \exp(\beta N)$$

with parameters $\alpha = 0.032 \pm 0.006$ and $\beta = 0.713 \pm 0.007$ (see Fig. A.1). Memory is measured in kilobytes. Using this estimate, we see that the maximum size we can reach with 1 Petabytes[1] is $N = 43$. Later on we will describe how to relax these requirements using parallelization.

3. **Propagation cost (time).** At the t-th propagation step, each of the $\binom{N}{t+1}$ entries of the vector $|x_{t+1}\rangle$ are computed by summing the contributions of $t+1$ vertices from the previous propagation step, *e.g.*

$$\langle j|x_{t+1}\rangle = w(i_1 \to j)\langle i_1|x_t\rangle + \cdots w(i_{t+1} \to j)\langle i_{t+1}|x_t\rangle$$

where $w(i \to j) = \Gamma/(\epsilon_x - \epsilon_j)$. We assume that ϵ_j can be computed using $O(1)$ FLOPs from the energy of a previous configuration, so that the computation of each weight $w(i \to j)$ takes $O(1)$ FLOPs. Then each entry takes $O(N)$ FLOPs. Therefore the t-th step of computation takes $O(N\binom{N}{t+1})$ FLOPs and the entire propagation on the hypercube of dimension N takes order of

$$\sum_{t=0}^{N-1} N\binom{N}{t+1} = N(2^N - 1) = O(N2^N)$$

floating-point operations. Again, this is confirmed by the data from the simulations we did, where a fit

$$\mathrm{Time}(N) = \alpha N \exp(\beta N)$$

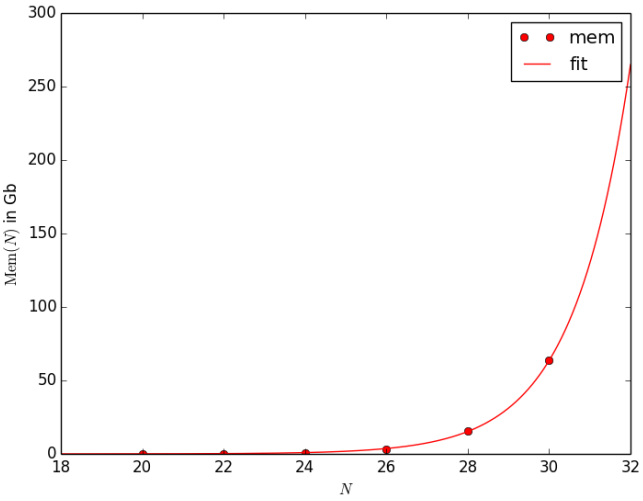


Figure A.1: fit for the memory usage of the algorithm for the full-propagation over the Boolean hypercube $\mathbb{B}^N = \{0,1\}^N$ (all N propagation steps), as a function of the system size N. Memory scale is shown in Megabytes. The curve $\alpha \exp(\beta N)$ obtained from the analytic estimate gives fit parameters $\alpha = 0.032 \pm 0.006$ and $\beta = 0.713 \pm 0.007$

gives fit parameters $\alpha = 0.032 \pm 0.006$ and $\beta = 0.713 \pm 0.007$ (see Fig. A.2). Memory is measured in seconds.

4. **Divide and conquer, parallelization.** In order to obtain the results presented in Chapter 4, the previous optimization was found to be sufficient. However, if one wants to access large system sizes, then one has to find a way around the memory requirements described above. One such way is parallelization: since the propagation process is done by a transfer matrix, *i.e.* a linear operator, then it is easy to split into independent subprocesses. As an example, suppose we want to map $|x_0\rangle$ to $|x_k\rangle = T^k|x_0\rangle$. One way of doing this would just be to apply the transfer matrix k times to the initial vector $|x_0\rangle$. However, one can easily see that the same result is obtained if we propagate $|x_0\rangle$ for a smaller time (say $k' \equiv k/2$) then expand the resulting vector $|x_{k/2}\rangle$ in the standard basis $\{|i\rangle\}_i$ and propagate each component independently by another $T^{k/2}$, summing the resulting vectors:

$$|x_k\rangle = T^k|x_0\rangle = T^{k/2}T^{k/2}|x_0\rangle = T^{k/2}|x_{k/2}\rangle \quad \text{(A.3)}$$

$$= T^{k/2}\Big(\sum_i \alpha_i|i\rangle\Big) = \sum_i \alpha_i T^{k/2}|i\rangle. \quad \text{(A.4)}$$

Of course one is not required to split the vector $|x_{k/2}\rangle$ into basis vectors. One might decide to use any decomposition $|x_{k/2}\rangle = |v_1\rangle + \cdots + |v_m\rangle$.

[1] approximately the memory available to the largest supercomputers available today

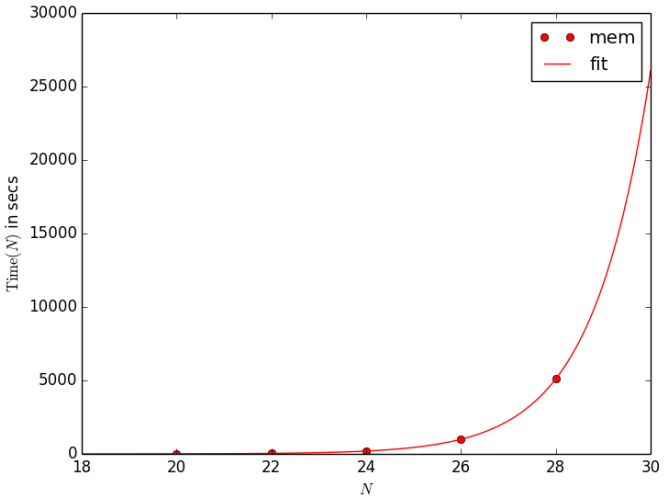


Figure A.2: fit for the time requirements of the algorithm for the full-propagation over the Boolean hypercube $\mathbb{B}^N = \{0,1\}^N$ (all N propagation steps), as a function of the system size N. Time scale is shown in seconds. The curve $\alpha N \exp(\beta N)$ obtained from the analytic estimate gives fit parameters $\alpha = 0.032 \pm 0.006$ and $\beta = 0.713 \pm 0.007$

We propose the following approach to bypass the memory bottleneck described above. The system is propagated up to a large enough time $t < N/2$ so that the vector $|x_t\rangle$ with $\binom{N}{t}$ entries is reasonably easy to store in memory. Then the goal is to propagate this vector to $|x_{N-t}\rangle$, a vector with other $\binom{N}{t}$ entries on the opposite side of the Boolean hypercube (*i.e.* on the other side of the memory bottleneck). This is done by expanding $|x_t\rangle$ in the computational basis

$$|x_t\rangle = \sum_i \alpha_i |i\rangle \tag{A.5}$$

(note that this is a sum of at most $\binom{N}{t}$ non-zero terms). Then each state $\alpha_i|i\rangle$ is propagated by $N-2t$ steps, giving a contribution to $|x_{N-t}\rangle$:

$$|x_{N-t}\rangle = T^{N-2t}|x_t\rangle = \sum_i \alpha_i T^{N-2t}|i\rangle$$

We will call each smaller propagation $\alpha_i|i\rangle \mapsto \alpha_i T^{N-2t}|i\rangle$ a "propagation-bridge". Each propagation bridge is equivalent to a propagation on a reduced Boolean hypercube $\mathbb{B}^M$ with $M = N-t$. The advantage with this approach is that the propagation-bridges need not be run simultaneously and the hard limit for the memory used is given by 1) the size of the two vectors $|x_t\rangle, |x_{N-t}\rangle$, and 2) the bottleneck of the propagation on the reduced hypercube, $\min\left\{\binom{N-t+1}{(N-t+1)/2}, \binom{N-t+1}{N-2t+1}\right\}$. The bottleneck of this

"bridge" method is of order

$$\mathrm{Mem}(N,t) = 2\binom{N}{t} + \min\left\{\binom{N-t+1}{(N-t+1)/2}, \binom{N-t+1}{N-2t+1}\right\}. \quad \text{(A.6)}$$

Fig. A.3 shows a comparison of the bottlenecks for the usual, full-propagation method and the bridge method. After we have reconstructed $|x_{N-t}\rangle$ we are past the memory bottleneck and the propagation can proceed again as usual.

Now, if we blindly apply the method just described one sees that the number of propagation-bridges one needs to compute in order to propagate $|x_t\rangle$ to $|x_{N-t}\rangle$ is very large, being $\binom{N}{t}$. The solution to this would be to expand $|x_t\rangle$ into a fixed number of vectors $|x_t\rangle = \sum_j |y_j^t\rangle$ rather than in the computational basis. One could for example collect the 2^N basis vectors $\{|i\rangle\}$ into r sets $S_1, \ldots, S_r$ and define (compare with Eq. (A.5))

$$|y_j^t\rangle \equiv \sum_{i \in S_j} \alpha_i |i\rangle.$$

Then one would need to propagate $|y_j^t\rangle$ to $T^{N-2t}|y_j^t\rangle$, for all $j = 1, \ldots, r$. This, depending on the choice of r, can be a much smaller number that $\binom{N}{t}$. The downside is of course that each of these larger propagation-bridges requires more memory than the ones described before. The optimal course will require to balance time and memory requirements.

A.2 PIMC

The path-integral Monte Carlo (PIMC) method applies the path-integral representation for the partition function of a quantum system in order to compute the thermal average $\langle \hat{O} \rangle \propto \mathrm{Tr}(\hat{O} e^{-\beta \hat{H}})$ of an observable $\hat{O}$ [93, 94]. In this way one identifies the thermodynamical properties of a d-dimensional quantum system with those of a $(d+1)$-dimensional classical system. The distribution of the classical observable is then sampled with a standard Markov-chain Monte Carlo. In order to be able to do this it is crucial that the weights that the quantum action associates to the paths be renormalizable to a *bona fide* probability distribution. This effectively means that they all need to be real and non-negative. The presence of weights with a negative sign completely invalidates this method, and one speaks of the emergence of a *sign problem*. In this appendix we review the derivation of the effective classical Hamiltonian for a transverse-field Ising spin glass system such as the one in Eq. (3.1). The same quantum Monte Carlo approach has also been employed in simulations of the quantum annealing of quantum Ising glasses [95, 12], and in the calculation of minimum gaps along adiabatic annealing dynamics [96].

Thanks to this quantum-to-classical mapping, the static properties of the system, including the thermodynamic potentials and the correlation functions, can be computed via standard Monte Carlo techniques by sampling (e.g., using the Metropolis algorithm) configurations according to the probability density function obtained from the derivation.

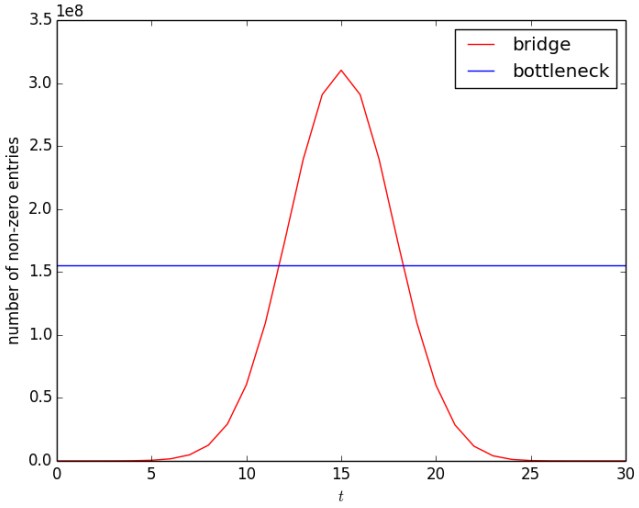


Figure A.3: Comparison of the memory bottlenecks for the full propagation method vs the bridge method, for a system of size $N = 30$. The full-propagation memory bottleneck has no t-dependence since its bottleneck is always given by $\binom{N+1}{(N+1)/2}$. The memory bottleneck of bridge method depends on the time t where we start breaking the propagation into smaller ones, as shown in Eq. (A.6).

Note that this derivation makes no assumption on the topology of the interaction graph of the spin-glass term. In this respect, this is completely general.

Note that the $e^{-\beta H}$ factor of a thermal state $\rho(\beta)$ is an imaginary-time propagator $U(-i\beta)$ for the system described by the Hamiltonian H. The first thing one needs to do is to split the imaginary-time interval β into m subintervals $\Delta\beta \equiv \beta/m$. Then the propagator $U(-i\beta)$ factorizes into

$$U(-i\beta) = U(-i\Delta\beta)\cdots U(-i\Delta\beta),$$

and the quantum partition function $Z \equiv \mathrm{Tr}(e^{-\beta H})$ can then be written as

$$\begin{aligned}\mathrm{Tr}(e^{-\beta H}) &= \mathrm{Tr}\left(\prod_{i=1}^{m} e^{-\Delta\beta H}\right) \\ &= \sum_{s^{(0)}}\cdots\sum_{s^{(m-1)}}\prod_{k=0}^{m-1}\langle s^{(k)}|e^{-\Delta\beta H}|s^{(k+1)}\rangle,\end{aligned}$$

where the upper indices (k) are taken modulo m, *i.e.* $(m) = (0)$. In the second line we have introduced $m-1$ resolutions of the identity $\mathbb{I} = \sum_{s^{(k)}} |s^{(k)}\rangle\langle s^{k}|$ between the $e^{-\Delta\beta H}$ factors. The outermost sum $\sum_{s^{(0)}}\langle s^{(0)}|\cdots|s^{(0)}\rangle$ is given by the trace operator. The choice for the basis vectors of these sums is, at this point, completely free. Nevertheless, we choose the product σ^z-basis for all k:

$$|s^{(k)}\rangle = |s_1\rangle \otimes |s_2\rangle \otimes \cdots \otimes |s_N\rangle \qquad \text{where } s_i \in \{\pm 1\} \text{ for all } i,$$

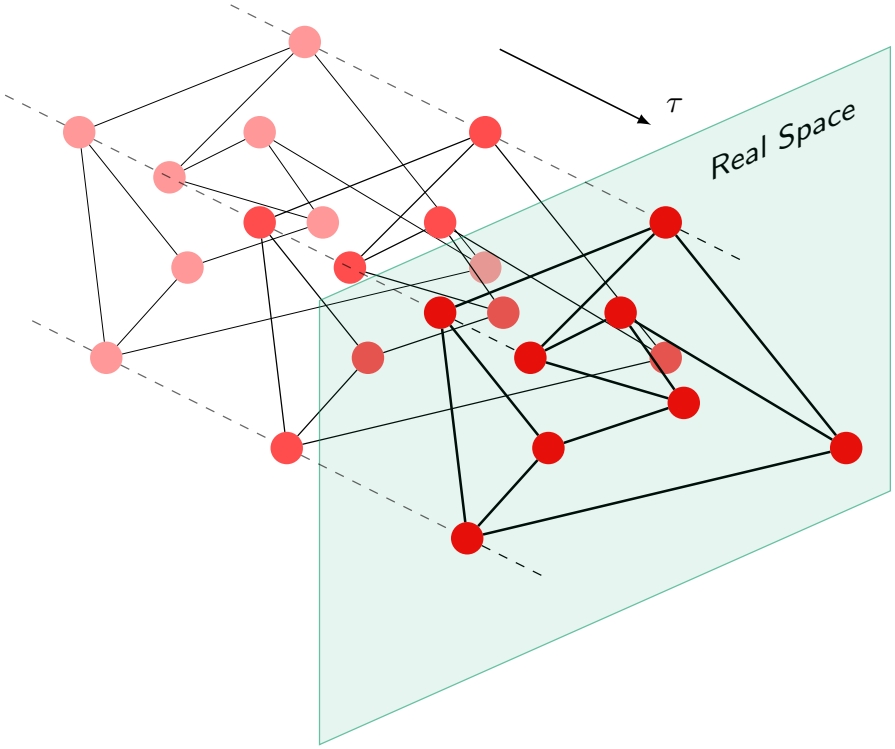


Figure A.4: A quantum theory on a cubic graph is turned via the quantum-to-classical mapping into an equivalent classical theory on a 5-regular graph, where an additional "imaginary time" dimension is to be taken into account. Effective couplings along this directions, some of which are represented by dashed lines in the above figure, replace the real-space transverse interactions of the quantum theory.

where we use the usual identification $|\uparrow\rangle \equiv |1\rangle, |\downarrow\rangle \equiv |-1\rangle$.

Next we compute the bra-ket term $\langle s^{(k)}|e^{-\Delta\beta H}|s^{(k+1)}\rangle$ where

$$H = -\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z - \Gamma\sum_i \sigma_i^x.$$

We apply the first-order approximation of the Suzuki-Trotter formula

$$e^{A+B} = \lim_{n\to\infty}\left(e^{A/n}e^{B/n}\right)^n$$

to the operator $e^{-\Delta\beta H}$ so that

$$\exp\left(\Delta\beta\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z + \Delta\beta\Gamma\sum_i\sigma_i^x\right) = \exp\left(\Delta\beta\sum_{\langle ij\rangle} J_{ij}\sigma_i^z\sigma_j^z\right) \times \exp\left(\Delta\beta\Gamma\sum_i\sigma_i^x\right) + O(\Delta\beta^2).$$

Note that the first factor on the r.h.s is diagonal in the chosen σ^z-basis, therefore in a term like $\langle s|e^{-\Delta\beta H}|s'\rangle$ we can multiply it to the left with the bra $\langle s|$ and obtain

$$\langle s|e^{-\Delta\beta H}|s'\rangle \approx e^{\Delta\beta\sum_{\langle ij\rangle} J_{ij}s_is_j}\langle s|e^{\Delta\beta\Gamma\sum_i\sigma_i^x}|s'\rangle,$$

where $s_i = \pm 1$ are the classical spin values that label the elements of the product state $|s\rangle = \bigotimes_i |s_i\rangle$. The approximation in the previous equation becomes exact in the limit of $m\to\infty$ (equivalently, when $\Delta\beta\to 0$).

Note that the energy term $\sum_i\sigma_i^x$ is non-interacting, in the sense that it is a sum of operators that act non-trivially on disjoint sets of spins. This means that we can write

$$\langle s|e^{\Delta\beta\Gamma\sum_i\sigma_i^x}|s'\rangle = \prod_{i=1}^N\langle s_i|e^{\Delta\beta\Gamma\sigma_i^x}|s_i'\rangle. \tag{A.7}$$

The last thing we need to do is to evaluate the single-spin bra-ket term $\langle s_i|e^{\Delta\beta\Gamma\sigma_i^x}|s_i'\rangle$. It is an exercise in trigonometry to prove that

$$\langle s_i|e^{\Delta\beta\Gamma\sigma_i^x}|s_i'\rangle = C\exp(Bs_is_i')$$

where

$$\begin{aligned} B &= -\frac{1}{2}\ln\tanh(\Delta\beta\Gamma) \\ C &= \sqrt{\frac{1}{2}\sinh(2\Delta\beta\Gamma)}. \end{aligned}$$

Then Eq. (A.7) becomes

$$C^N\exp\left(J^\perp\Delta\beta\sum_i J_{ij}S_i^{(k)}S_i^{(k+1)}\right)$$

where

$$J^\perp \equiv -\frac{1}{2\Delta\beta}\ln\tanh\left(\Gamma\Delta\beta\right),$$

and the partition function $Z = \mathrm{Tr}(e^{-\beta H})$ of the quantum system can be approximated by

$$Z_m \equiv C^{Nm} \sum_{\vec{s}} e^{-\Delta\beta H_{\mathrm{eff}}}$$

where H_{eff} is the Hamiltonian of a system of Nm *classical* spins:

$$H_{\mathrm{eff}}(\vec{s}) \equiv \sum_{k=0}^{m-1} \Big(- \sum_{\langle i,j \rangle} J_{ij} S_i^{(k)} S_j^{(k)} - J^{\perp} \sum_i S_i^{(k)} S_i^{(k+1)} \Big).$$

This is a system that can be easily simulated using standard Monte Carlo techniques.

A.3 The Trotterization of Replica Systems

The computation of entropies, and in particular of the Rényi entanglement entropy, is a challenging computational task, even when the Monte Carlo simulations of the effective classical model can be efficiently performed without negative sign problems nor frustration. Several approaches to compute Rényi entropies using Monte Carlo simulations have been engineered, including the temperature-integration method of Ref. [97], the swap operator method of Ref. [98] (for $SU(2)$-invariant lattice spin systems), the weight-ratio estimator method of Ref. [99, 100, 101], the mixed-ensemble method of Ref. [102], and the extended configuration-space method of Ref [58]. While all these methods have their own appealing features, we chose to adopt the extended configuration-space method, since it is extremely versatile and, more importantly, efficient in the large system-size regime. We briefly describe it below, following Ref [58].

This method is based on Eq. (1.21) of Chapter 1, derived in Ref. [41]:

$$\mathrm{Tr}(\rho_A^2) = \sum_{a,a'} \langle a|\rho_A|a'\rangle\langle a'|\rho_A|a\rangle = \frac{Z_A^{(2)}}{Z^2}$$

where $Z_A^{(2)}$ is the "partition function" of two replicas of the systems glued along the region A in the imaginary-time direction

$$Z_A^{(2)} = \sum_{a,a',b,b'} \langle a|\langle b|e^{-\beta H}|a'\rangle|b\rangle\langle a'|\langle b'|e^{-\beta H}|a\rangle|b'\rangle,$$

and Z^2 is the partition function of two independent replicas

$$Z^2 = \sum_{a,a',b,b'} \langle a|\langle b|e^{-\beta H}|a\rangle|b\rangle\langle a'|\langle b'|e^{-\beta H}|a'\rangle|b'\rangle.$$

We can apply the same path-integral mapping described in the previous section to both of these partition functions, obtaining two classical Hamiltonians (respectively, H_A and $H_\emptyset$) defined on the same system of $2Nm$ classical spins. The form of the Hamiltonian is the same for both cases $X \in \{A, \emptyset\}$

$$H_X \equiv \sum_{k=0}^{2m-1} \Big(- \sum_{\langle i,j \rangle} J_{ij} S_i^{(k)} S_j^{(k)} - J^{\perp} \sum_{i\in A} S_i^{(k)} \tilde{S}_i^{(k+1)} - J^{\perp} \sum_{i\in B} S_i^{(k)} \tilde{S}_i^{(k+1)} \Big).$$

where $\{S_i^{(k)} \mid 1 \leq i \leq N, 0 \leq k \leq 2m-1\}$ is the set of spins that composes the system, and the notation $\tilde{S}_i^{(k)}$ is used to explain the difference between the two cases in the boundary conditions in the imaginary-time direction. For the independent-replicas Hamiltonian $H_\emptyset$, all imaginary-time loops close on themselves after an imaginary-time interval β, for both replicas. This means that the spins $\tilde{S}_i^{(k)}$ satisfy, for all $i = 1, \dots N$

- $\tilde{S}_i^{(k)} = S_i^{(k)}$ for all "bulk" timeslices $k \neq m, 2m$
- $\tilde{S}_i^{(m)} = S_i^{(0)}$ and $\tilde{S}_i^{(2m)} = S_i^{(m)}$ for the two "boundary" timeslices $m, 2m$.

For the glued-replicas Hamiltonian H_A, if i is a spin in the region B, then the imaginary-time loops close on themselves after an imaginary-time interval β, the same as before. Instead, if $i \in A$, then the imaginary-time loops close on themselves after an imaginary-time interval 2β. The boundary conditions on these spins are

- $\tilde{S}_i^{(k)} = S_i^{(k)}$ for all "bulk" timeslices $k \neq 2m$
- $\tilde{S}_i^{(2m)} = S_i^{(0)}$ for the single "boundary" timeslice $2m$.

These boundary conditions are shown in Fig. A.5.

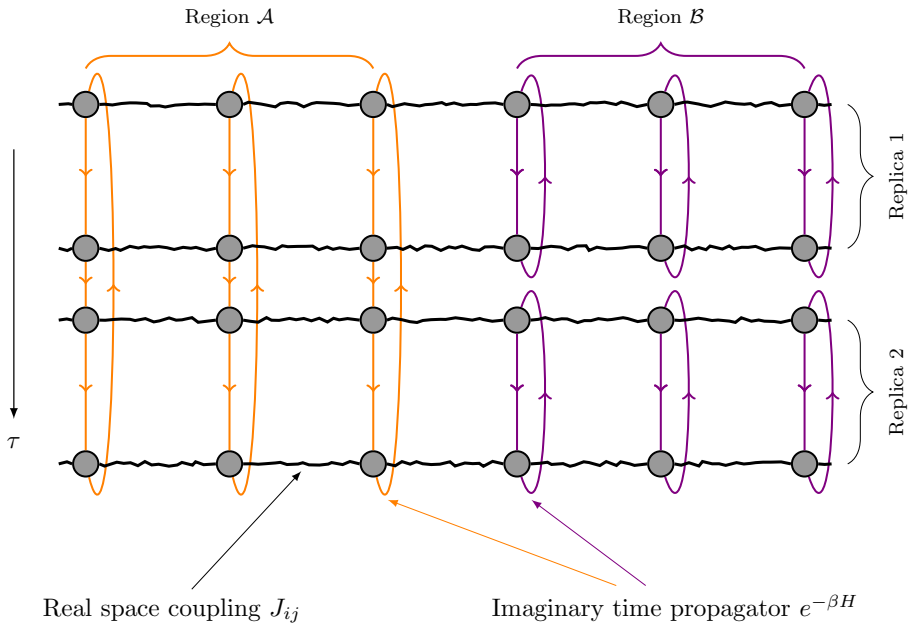


Figure A.5: Pictorial representation of the system described by $Z_A^{(2)}$. Two copies of the original system are connected by imaginary time couplings, but only in the region A. Wobbly lines denote random interactions, solid lines are ferromagnetic.

Now, for each classical spin configuration $x \in \{\pm 1\}^{2Nm}$ we can define two weights

$$w_\emptyset(x) \equiv e^{-\Delta\beta H_\emptyset(x)}, \qquad w_A(x) \equiv e^{-\Delta\beta H_A(x)}$$

A.4 Humeniuk-Roscilde Algorithm

In [58] the replica method was recast as a quantum Monte Carlo algorithm capable in principle to compute the Rényi entropies of generic systems. As we think the full proof is instructive and does not – to our knowledge – appear in the literature, we describe it here.

If X is the configuration space of the Trotterized system (*i.e.* the set of all possible classical spin states), define an extended state space $S = X \times \{\emptyset, A\}$. A state in S is a pair (x, i) where $x \in X$ is a configuration of the original system and $i \in \emptyset, A$ is a label for one of the two partition functions $Z_A^{(2)}$ and Z^2. Define over S the (possibli unnormalized) probability distribution

$$\tilde{P}(x,i) = \begin{cases} w(x) & \text{if } i = \emptyset \\ w_A(x) & \text{if } i = A, \end{cases}$$

where $w(x)$ is the weight that the partition function Z^2 gives to the configuration x and $w_A(x)$ is the weight that the partition function $Z_A^{(2)}$ gives to the same configuration. Call $P \equiv \frac{1}{\mathcal{N}}\tilde{P}$ its normalized version. Note that normalizing P over the configurational veriable x we have

$$P(A) \equiv \frac{1}{\mathcal{N}} \sum_{x \in X} \tilde{P}(x, A) = \frac{1}{\mathcal{N}} \sum_{x \in X} w_A(x) = \frac{1}{\mathcal{N}} Z_A^{(2)}$$

and

$$P(\emptyset) \equiv \frac{1}{\mathcal{N}} \sum_{x \in X} \tilde{P}(x, \emptyset) = \frac{1}{\mathcal{N}} \sum_{x \in X} w(x) = \frac{1}{\mathcal{N}} Z^2.$$

Therefore the ratio of the two marginals give

$$\frac{P(A)}{P(\emptyset)} = \frac{\mathcal{N}}{\mathcal{N}} \frac{Z_A^{(2)}}{Z^2} = \frac{Z_A^{(2)}}{Z^2}.$$

The bottom line is that by sampling P and by keeping track of how many times we get a state of type (x, A) and $(x, \emptyset)$ we can learn the ratio of the two partition functions. Let $N_A(m)$ be the number of times we got a state of type (x, A) after m samplings, and let $N_\emptyset(m)$ be the number of times we got a state of type $(x, \emptyset)$. Then for $m \to \infty$ we have that $N_A(m)/N_\emptyset(m) \to P(A)/P(\emptyset)$. What is needed now is a way of sampling the probability P. We construct a Markov chain over S that has P as its limiting steady-state distribution.

A.4.1 Transition matrix

The transition matrix is a matrix $T(x, i \to x', i')$ that gives the probability of transitioning from a state $(x, i) \in S$ to a new state $(x', i') \in S$. Instead of explicitly defining its entries we describe it algorithmically. Fix a real number $0 < p < 1$. If we are in a state $(x, i) \in S$ we

- with probability p do the usual Metropolis-Hastings algorithm according to the probability $P_i(x) \equiv P(x,i)$:
  1. generate a new configuration x' according to a symmetric (*i.e.* $Q(x|x') = Q(x'|x)$) conditional probability $Q(x|x')$.
  2. accept the new configuration x' with probability
  $$\Pr[\text{accept } x'] = \min\left(1, \frac{\tilde{P}(x',i)}{\tilde{P}(x,i)}\right)$$
  3. if accepted, update the configuration part of the state $(x,i) \mapsto (x',i)$.
- with probability $1-p$ do the following
  1. generate a new dynamical label $j \in \{\emptyset, A\}$ according to a symmetric conditional probability $G(j|i)$. In Ref. [58] this probability is defined by $G(A|A) = G(\emptyset|\emptyset) = 0$ and $G(A|\emptyset) = G(\emptyset|A) = 1$, *i.e.* one always tries to move to the *other* dynamics.
  2. accept the new candidate j according to the probability
  $$\Pr[\text{accept } j] = \min\left(1, \frac{\tilde{P}(x,j)}{\tilde{P}(x,i)}\right)$$
  3. if accepted, update the dynamical part of the state $(x,i) \mapsto (x,j)$.

Note that this method is similar to the *simulated tempering* Monte Carlo technique [103], a modified version of simulated annealing that was proposed to more easily study rough energy landscapes. In simulated tempering the inverse temperature β is a dynamical variable in the simulation, here instead β is kept constant but the Hamiltonian of the replicated system is updated dynamically.

It is worth mentioning that the efficiency of this algorithm might degrade when the subsystem size l_A increases, due to the decrease of the acceptance rate of the Monte Carlo updates that switch between the two sectors Z^2 and $Z_A^{(2)}$, resulting in large statistical fluctuations. This eventual problem could be circumvented by using the incremental formula implemented in Refs. [97, 58], where the Rényi entropy of a (large) subsystem A is obtained by decomposing it into subparts of increasing size; however, we verified that in the disordered RRG considered in this work, the basic algorithm (without the incremental decomposition) is sufficiently efficient even for considerably large system sizes, and that the incremental formula does not provide a critical performance boost.

A.4.2 Detailed Balance

We prove that the probability distribution $P(x,i)$ and the transition matrix $T(x, i \to x', i')$ implicitly defined by the algorithm of the previous section satisfy detailed balance, that is

$$P(x,i)T(x, i \to x', i') = P(x',i')T(x', i' \to x, i) \tag{A.8}$$

for all $x, x' \in X$ and $i, i' \in \{\emptyset, A\}$. This entails that the Markov chain defined by T has P as a limiting distribution. We distinguish four cases, for all possible identity relations between the elements x, x' and i, i' in Eq. (A.8).

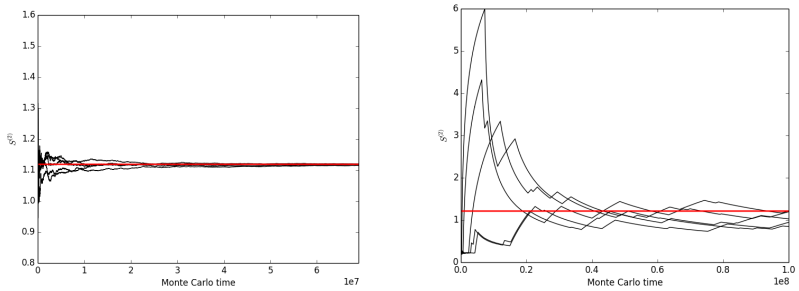


Figure A.6: Convergence times of the Humeniuk-Roscilde algorithm for the computation of the Rényi entanglement entropy. Shown here is the cumulative estimator $-\log(N_A/N_\emptyset)$ for $S^{(2)}$. The black lines represent simulations of the same realization of disorder initialized with different initial states. The red line is the exact value obtained by exact diagonalization. The left Figure shows simulations at $\Gamma = 1.5$. The evolution is essentially ergodic at this system size. In the right Figure (at $\Gamma = 0.7$) the glassy dynamics translates to longer convergence times for the Monte Carlo simulation.

- $x = x'$ **and** $i = i'$ **case.** Trivial.

- $x \neq x'$ **and** $i \neq i'$ **case.** Trivial, as the transition $(x, i) \mapsto (x', i')$ has zero probability of happening.

- $x \neq x'$ **and** $i = i'$ **case.** This means that the configurational part of the state was updated. This happens with probability

$$p \cdot \min\left(1, \frac{\tilde{P}(x', i)Q(x|x')}{\tilde{P}(x, i)Q(x'|x)}\right)$$

  By defining $\tilde{P}_i(x) \equiv \tilde{P}(x, i)$ then the proof follows the textbook case for the standard Metropolis-Hastings algorithm.

- $x = x'$ **and** $i \neq i'$ **case.** This means that the dynamical part of the state was updated. This happens with probability

$$(1 - p) \cdot \min\left(1, \frac{\tilde{P}(x, i')G(i'|i)}{\tilde{P}(x, i)G(i|i')}\right).$$

Now define $k \in \mathbb{R}$ as

$$k \equiv \frac{\tilde{P}(x, i')G(i'|i)}{\tilde{P}(x, i)G(i|i')}.$$

Suppose $k \geq 1$ then $1/k \leq 1$. Then

$$P(x, i)T(x, i \to x, i') = (1 - p)P(x, i)$$

and

$$\begin{aligned} P(x,i')T(x,i' \to x,i) &= (1-p)P(x,i')\frac{\tilde{P}(x,i)G(i|i')}{\tilde{P}(x,i')G(i'|i)} \\ &= (1-p)\frac{1}{\mathcal{N}}\tilde{P}(x,i) \\ &= (1-p)P(x,i) \end{aligned}$$

and therefore detailed balance is satisfied. The case where $k \leq 1$ is perfectly analogous.

A.5 Simulated Annealing

Simulated annealing is a numerical technique designed to find the solution to combinatorial optimization problems. Given a COP $f(X_1, \ldots, X_N) = \sum_i f_i(X_1, \ldots, X_N)$ over N Boolean variables $X_1, \ldots, X_N$, one defines N classical spins $S_i \equiv -2X_i + 1$. Then a Hamiltonian H is defined over this fictitious spins

$$H(S_1, \ldots, S_N) = E + \sum_i h_i S_i + \sum J_{ij} S_i S_j + \sum_{i,j,k} J_{ijk} S_i S_j S_k + \cdots$$

so that the $f(X_1, \ldots, X_N) = H(S_1, \ldots, S_N)$ and therefore the ground state of H corresponds to the solution to the minimization problem f.

Then the thermal state of the fictitious spin system at temperature T is described by the Gibbs probability distribution

$$P_T(x) = \frac{1}{Z} e^{-H(x)/(Tk_B)}$$

where $Z = \sum_x e^{H(x)/(Tk_B)}$ is the partition function. At high temperature all the microstates are approximately equiprobable while at low temperature the probability distribution will concentrate on the low-energy states of H.

The idea is to implement numerically the quasi-static transformation that takes the system from an infinite-temperature state, where all microstates x are equiprobable, to a zero-temperature state where the system settles to one of the (possibly degenerate) ground states. This is usually done with a Markov Chain Monte Carlo, a randomized algorithm that is able to sample the Gibbs distribution of a system at a given temperature T starting from any given initial state.

A common run of simulated annealing goes as follows. At the beginning of the algorithm, a state x is generated uniformly at random. By definition, this state is typical of a $T = \infty$ thermal state. Then the system is thermalized to a high enough temperature T_∞ so that the Gibbs probability P_{T_∞} can be reasonably taken to be a good approximation to the uniform probability distribution over X. At successive steps, the temperature is lowered by a finite amount ΔT and the system is allowed to thermalize at the new temperature. By repeatedly lowering the temperature, $T = 0$ is reached eventually.

Special care must be taken when the system enters a glassy phase at low temperature, as is often the case for combinatorial optimization problems. The issue in this case is that, due to the emergence of a rough free energy landscape at low temperatures, if in the simulation the temperature is changed too abruptly

(ΔT is too large) or if these changes are too close together in time (so the system is not allowed to thermalize at the new, lower temperature), then the system will go out of equilibrium, it will get stuck in a metastable state and will not reach the ground state at the end of the algorithm.

Bibliography

[1] Gianni Mossi, Tommaso Parolini, Sebastiano Pilati, and Antonello Scardicchio. On the quantum spin glass transition on the bethe lattice. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(1):013102, 2017.

[2] Gianni Mossi and Antonello Scardicchio. Ergodic and localized regions in quantum spin glasses on the bethe lattice. *arXiv:1703.03678*, 2017.

[3] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, Jun 1982.

[4] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Phys. Rev. E*, 58:5355–5363, Nov 1998.

[5] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001.

[6] Dorit Aharonov, Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, and Oded Regev. Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM Rev.*, 50(4):755–787, November 2008.

[7] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, STOC '71, pages 151–158, New York, NY, USA, 1971. ACM.

[8] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001.

[9] Marko Žnidarič. Scaling of the running time of the quantum adiabatic algorithm for propositional satisfiability. *Phys. Rev. A*, 71:062305, Jun 2005.

[10] Itay Hen and A. P. Young. Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems. *Phys. Rev. E*, 84:061152, Dec 2011.

[11] Marko Žnidarič and Martin Horvat. Exponential complexity of an adiabatic algorithm for an np-complete problem. *Phys. Rev. A*, 73:022329, Feb 2006.

[12] Roman Martoňák, Giuseppe E Santoro, and Erio Tosatti. Quantum annealing by the path-integral monte carlo method: The two-dimensional random ising model. *Physical Review B*, 66(9):094203, 2002.

[13] Demian A. Battaglia, Giuseppe E. Santoro, and Erio Tosatti. Optimization by quantum annealing: Lessons from hard satisfiability problems. *Phys. Rev. E*, 71:066707, Jun 2005.

[14] Bettina Heim, Troels F. Rønnow, Sergei V. Isakov, and Matthias Troyer. Quantum versus classical annealing of ising spin glasses. *Science*, 348(6231):215–217, 2015.

[15] Sergio Boixo, Troels F Rønnow, Sergei V Isakov, Zhihui Wang, David Wecker, Daniel A Lidar, John M Martinis, and Matthias Troyer. Evidence for quantum annealing with more than one hundred qubits. *Nature Physics*, 10(3):218–224, 2014.

[16] Alexandre M. Zagoskin, Evgeni Il'ichev, Miroslav Grajcar, Joseph J. Betouras, and Franco Nori. How to test the "quantumness" of a quantum computer? *Frontiers in Physics*, 2(33), 2014.

[17] Sergio Boixo, Vadim N. Smelyanskiy, Alireza Shabani, Sergei V. Isakov, Mark Dykman, Vasil S. Denchev, Mohammad Amin, Anatoly Smirnov, Masoud Mohseni, and Hartmut Neven. Computational role of collective tunneling in a quantum annealer, 2014.

[18] Seung Woo Shin, Graeme Smith, John A. Smolin, and Umesh Vazirani. How "quantum" is the d-wave machine?, 2014.

[19] Steven H. Adachi and Maxwell P. Henderson. Application of quantum annealing to training of deep neural networks. *arXiv:1510.06356*, 2015.

[20] Florian Neukart, Gabriele Compostella, Christian Seidel, David von Dollen, Sheir Yarkoni, and Bob Parney. Traffic flow optimization using a quantum annealer. *arXiv:1708.01625*, 2017.

[21] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *arXiv:1611.09347*, 2016.

[22] A. Messiah. *Quantum Mechanics*. Dover books on physics. Dover Publications, 1961.

[23] Andris Ambainis and Oded Regev. An elementary proof of the quantum adiabatic theorem. *arXiv preprint arXiv:quant-ph/0411152*, 2006.

[24] M. Born and V. Fock. Beweis des adiabatensatzes. *Zeitschrift für Physik*, 51(3):165–180, Mar 1928.

[25] Giuseppe Santoro. Non-equilibrium quantum systems (Lecture notes). *http://indico.ictp.it/event/7644/session/2/contribution/5/material/2/0.pdf*, 2016.

[26] Sabine Jansen, Mary-Beth Ruskai, and Ruedi Seiler. Bounds for the adiabatic approximation with applications to quantum computation. *Journal of Mathematical Physics*, 48(10):102111, 2007.

[27] Donny Cheung, Peter Høyer, and Nathan Wiebe. Improved error bounds for the adiabatic approximation. *Journal of Physics A: Mathematical and Theoretical*, 44(41):415302, 2011.

[28] Satoshi Morita and Hidetoshi Nishimori. Mathematical foundation of quantum annealing. *Journal of Mathematical Physics*, 49(12):125210, 2008.

[29] Lov Grover. A fast quantum mechanical algorithm for database search. *arXiv preprint arXiv:quant-ph/9605043*, 1996.

[30] Lev D Landau. Zur theorie der energieubertragung. ii. *Phys. Z. Sowjetunion*, 2(46):1–13, 1932.

[31] Clarence Zener. Non-adiabatic crossing of energy levels. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 137(833):696–702, 1932.

[32] R. MacKenzie, E. Marcotte, and H. Paquette. Perturbative approach to the adiabatic approximation. *Phys. Rev. A*, 73:042104, Apr 2006.

[33] Bela Bauer, Lei Wang, Iztok Pižorn, and Matthias Troyer. Entanglement as a resource in adiabatic quantum optimization. *arXiv preprint arXiv:1501.06914v1*, 2015.

[34] P. J. D. Crowley, T. Đurić, W. Vinci, P. A. Warburton, and A. G. Green. Quantum and classical dynamics in adiabatic computation. *Phys. Rev. A*, 90:042317, Oct 2014.

[35] Luca Pezzé and Augusto Smerzi. Quantum theory of phase estimation. *arXiv preprint arXiv:1411.5164*, 2014.

[36] Philipp Hyllus, Wiesław Laskowski, Roland Krischek, Christian Schwemmer, Witlef Wieczorek, Harald Weinfurter, Luca Pezzé, and Augusto Smerzi. Fisher information and multiparticle entanglement. *Phys. Rev. A*, 85:022321, Feb 2012.

[37] Géza Tóth. Multipartite entanglement and high-precision metrology. *Phys. Rev. A*, 85:022322, Feb 2012.

[38] Román Orús and José I. Latorre. Universality of entanglement and quantum-computation complexity. *Phys. Rev. A*, 69:052308, May 2004.

[39] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91:147902, Oct 2003.

[40] Richard Jozsa and Noah Linden. On the role of entanglement in quantum-computational speed-up. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 459(2036):2011–2032, 2003.

[41] Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(06):P06002, 2004.

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

[43] H. A. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 150(871):552–575, 1935.

[44] Nicholas C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *Journal of Combinatorial Theory, Series B*, 31(2):168 – 182, 1981.

[45] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:13 pp., 2001.

[46] Miklos Abért, Andreas Thom, and Balint Virág. Benjamini-schramm converegence and pointwise convergence of the spectral measure. *preprint available on http://www.math.uni-leipzig.de/MI/thom/*, 2013.

[47] Doron Puder. Expansion of random graphs: new proofs, new results. *Inventiones mathematicae*, 201(3):845–908, Sep 2015.

[48] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[49] Marc Mézard and Giorgio Parisi. The bethe lattice spin glass revisited. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(2):217–233, 2001.

[50] DJ Thouless. Spin-glass on a bethe lattice. *Physical review letters*, 56(10):1082, 1986.

[51] JRL De Almeida and David J Thouless. Stability of the sherrington-kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983, 1978.

[52] F Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

[53] Mark S. Pinsker. On the complexity of a concentrator. In *7th International Teletraffic Conference*, 1973.

[54] C Laumann, A Scardicchio, and SL Sondhi. Cavity method for quantum spin glasses on the bethe lattice. *Physical Review B*, 78(13):134424, 2008.

[55] Edward Farhi, David Gosset, Itay Hen, AW Sandvik, Peter Shor, AP Young, and Francesco Zamponi. Performance of the quantum adiabatic algorithm on random instances of two optimization problems on regular hypergraphs. *Physical Review A*, 86(5):052334, 2012.

[56] CR Laumann, A Pal, and A Scardicchio. Many-body mobility edge in a mean-field quantum spin glass. *Physical review letters*, 113(20):200405, 2014.

[57] CL Baldwin, CR Laumann, A Pal, and A Scardicchio. The many-body localized phase of the quantum random energy model. *Physical Review B*, 93(2):024202, 2016.

[58] Stephan Humeniuk and Tommaso Roscilde. Quantum monte carlo calculation of entanglement rényi entropies for generic quantum systems. *Physical Review B*, 86(23):235116, 2012.

[59] A. Steger and N. C. Wormald. Generating random regular graphs quickly. *Combinatorics, Probability and Computing*, 8(4):377–396, 1999.

[60] Luigi Amico, Rosario Fazio, Andreas Osterloh, and Vlatko Vedral. Entanglement in many-body systems. *Rev. Mod. Phys.*, 80:517–576, May 2008.

[61] R. Abou-Chacra, DJ Thouless, and PW Anderson. A selfconsistent theory of localization. *Journal of Physics C: Solid State Physics*, 6:1734, 1973.

[62] Michael Aizenman and Simone Warzel. Resonant delocalization for random schrödinger operators on tree graphs. *J. Eur. Math. Soc.*, 15(4):1167–1222, 2013.

[63] P.W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109(5):1492–1505, 1958.

[64] AJ Bray and MA Moore. Replica theory of quantum spin glasses. *Journal of Physics C: Solid State Physics*, 13(24):L655, 1980.

[65] Yadin Y Goldschmidt and Pik-Yin Lai. Ising spin glass in a transverse field: Replica-symmetry-breaking solution. *Physical review letters*, 64(21):2467, 1990.

[66] P. Ray, B. K. Chakrabarti, and Arunava Chakrabarti. Sherrington-kirkpatrick model in a transverse field: Absence of replica symmetry breaking due to quantum fluctuations. *Phys. Rev. B*, 39:11828–11832, Jun 1989.

[67] B. Altshuler, H. Krovi, and J. Roland. Adiabatic quantum optimization fails for random instances of np-complete problems. *arXiv:0908.2782*, 2009.

[68] D. Ruelle. A remark on bound states in potential-scattering theory. *Il Nuovo Cimento A (1965-1970)*, 61(4):655–662, Jun 1969.

[69] W. O. Amrein and V. Georgescu. On the characterization of bound states and scattering states in quantum mechanics. 46(5):635–658, Jan 1974.

[70] Volker Enss. Asymptotic completeness for quantum mechanical potential scattering. i. short range potentials. *Comm. Math. Phys.*, 61(3):285–291, 1978.

[71] B.C. Hall. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics. Springer New York, 2013.

[72] M. Aizenman and S. Warzel. *Random Operators:*. Graduate Studies in Mathematics. American Mathematical Society, 2015.

[73] B. Altshuler, H. Krovi, and J. Roland. Anderson localization casts clouds over adiabatic quantum optimization. *arxiv:0912.0746*, 2009.

[74] B. Altshuler, H. Krovi, and J. Roland. Anderson localization makes adiabatic quantum optimization fail. *Proceedings of the National Academy of Sciences*, 107(28):12446, 2010.

[75] Sergey Knysh and Vadim Smelyanskiy. On the relevance of avoided crossings away from quantum critical point to the complexity of quantum adiabatic algorithm. *arXiv preprint arXiv:1005.3011*, 2010.

[76] Neil G. Dickson and M. H. S. Amin. Does adiabatic quantum optimization fail for np-complete problems? *Phys. Rev. Lett.*, 106:050502, Feb 2011.

[77] Edward Farhi, Jeffrey Goldstone, David Gosset, Sam Gutmann, Hervey B Meyer, and Peter Shor. Quantum Adiabatic Algorithms, Small Gaps, and Different Paths. *arXiv:0909.4766*, 2009.

[78] Francesca Pietracaprina, Valentina Ros, and Antonello Scardicchio. Forward approximation as a mean-field approximation for the anderson and many-body localization transitions. *Phys. Rev. B*, 93:054201, Feb 2016.

[79] G.A. Baker and P. Graves-Morris. *Padé Approximants*. Encyclopedia of Mathematics an. Cambridge University Press, 1996.

[80] DM Basko, IL Aleiner, and BL Altshuler. Metal–insulator transition in a weakly interacting many-electron system with localized single-particle states. *Annals of physics*, 321(5):1126–1205, 2006.

[81] John Z. Imbrie. On many-body localization for quantum spin chains. *Journal of Statistical Physics*, 163(5):998–1048, Jun 2016.

[82] John Z. Imbrie. Diagonalization and many-body localization for a disordered quantum spin chain. *Phys. Rev. Lett.*, 117:027201, Jul 2016.

[83] Eman Hamza, Robert Sims, and Günter Stolz. Dynamical localization in disordered quantum spin systems. *Communications in Mathematical Physics*, 315(1):215–239, Oct 2012.

[84] Oriol Bohigas and Marie-Joya Giannoni. *Chaotic motion and random matrix theories*, pages 1–99. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984.

[85] and. Level clustering in the regular spectrum. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 356(1686):375–394, 1977.

[86] E. B. Bogomolny, B. Georgeot, M.-J. Giannoni, and C. Schmit. Chaotic billiards generated by arithmetic groups. *Phys. Rev. Lett.*, 69:1477–1480, Sep 1992.

[87] Akhilesh Pandey and Ramakrishna Ramaswamy. Level spacings for harmonic-oscillator systems. *Phys. Rev. A*, 43:4237–4243, Apr 1991.

[88] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, 65(3):239–362, 2016.

[89] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Phys. Rev. B*, 75:155111, Apr 2007.

[90] W. van Dam, M. Mosca, and U. Vazirani. How powerful is adiabatic quantum computation? In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 279–287, Oct 2001.

[91] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi. The quantum adiabatic algorithm applied to random optimization problems: The quantum spin glass perspective. *Physics Reports*, 523(3):127 – 205, 2013. The Quantum Adiabatic Algorithm Applied to Random Optimization Problems: The Quantum Spin Glass Perspective.

[92] M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford Graduate Texts. OUP Oxford, 2009.

[93] Masuo Suzuki. Relationship between d-dimensional quantal spin systems and (d+ 1)-dimensional ising systems equivalence, critical exponents and systematic approximants of the partition function and spin correlations. *Progress of Theoretical Physics*, 56(5):1454–1469, 1976.

[94] Masuo Suzuki. *Quantum Monte Carlo Methods in Equilibrium and Nonequilibrium Systems: Proceedings of the Ninth Taniguchi International Symposium, Susono, Japan, November 14–18, 1986*, volume 74. Springer Science & Business Media, 2012.

[95] Giuseppe E Santoro, Roman Martoňák, Erio Tosatti, and Roberto Car. Theory of quantum annealing of an ising spin glass. *Science*, 295(5564):2427–2430, 2002.

[96] A Peter Young, Sergey Knysh, and Vadim N Smelyanskiy. Size dependence of the minimum excitation gap in the quantum adiabatic algorithm. *Physical review letters*, 101(17):170503, 2008.

[97] Roger G. Melko, Ann B. Kallin, and Matthew B. Hastings. Finite-size scaling of mutual information in monte carlo simulations: Application to the spin-$\frac{1}{2}$ xxz model. *Phys. Rev. B*, 82:100409, Sep 2010.

[98] Matthew B Hastings, Iván González, Ann B Kallin, and Roger G Melko. Measuring renyi entanglement entropy in quantum monte carlo simulations. *Physical review letters*, 104(15):157201, 2010.

[99] Michele Caraglio and Ferdinando Gliozzi. Entanglement entropy and twist fields. *Journal of High Energy Physics*, 2008(11):076, 2008.

[100] Ferdinando Gliozzi and Luca Tagliacozzo. Entanglement entropy and the complex plane of replicas. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(01):P01002, 2010.

[101] Vincenzo Alba, Luca Tagliacozzo, and Pasquale Calabrese. Entanglement entropy of two disjoint intervals in c = 1 theories. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(06):P06012, 2011.

[102] P.V. Buividovich and M.I. Polikarpov. Numerical study of entanglement entropy in lattice gauge theory. *Nuclear Physics B*, 802(3):458 – 474, 2008.

[103] E. Marinari and G. Parisi. Simulated tempering: A new monte carlo scheme. *EPL (Europhysics Letters)*, 19(6):451, 1992.