博 士 論 文

Quantum Annealing: Quantum Statistical Physics applied to Classical Computational Hardness

（量子アニーリング：

古典的計算困難性への量子統計物理の応用）

高橋 惇

Contents

Chapter 1

Introduction

The most important goal of physics is to **understand** *the physical world.* While this statement may be regarded as a common claim, it is rather nontrivial what we exactly mean by an *understanding.* There may be many different definitions of understanding, but there are conditions we would certainly hope for when someone claims to understand something (say, a system). Will she be able to *predict* how the system will behave? While it may be too much to require her to predict *everything* about the system, it will certainly be daft to claim she understands the system even if she can't predict *anything* about it. The ultimate form of prediction is a *simulation* of the system, which all possible predictions could be derived from it. Therefore, it seems reasonable to say that the amount/quality of the prediction/simulation someone could have for a system reflects how much understanding she has about the system[*1].

It is therefore crucial to know how far we can simulate a given physical system, and how. A ubiquitous phenomenon regarding this point, central in physics, is the *hardness* of simulation, or perhaps hardness of problems in general. Originally dating back to the three-body problem [4], it was known that it is impossible to write down analytically the fate of more than three interacting mass points. Obviously, this does not immediately imply that it is impossible to simulate since various numerical methods have been developed. However, for people back then, writing down the analytic form of trajectories was almost equivalent to simulating the trajectories, since there were no computers. Even today, it still remains a hard problem, in the sense that it requires an exponential precision of the initial configuration due to its chaotic behavior.

Hardness of simulation does not always become an obstruction, and sometimes plays the role of a starting point. For example, since it is already hard to simulate the motion of three particles, it is *extremely* hard, in fact, may be regarded as impossible, to know all of the trajectories of 10^{23} particles which we face in everyday life. Nevertheless, it is this impossibility which actually enables us to use the framework of statistical mechanics, a framework known to produce extremely accurate predictions. It could be seen as the hardness plays a certain role in actually deepening our understandings of macroscopic systems.

Hardness also arises naturally when we consider problems regarding physical systems. For example, given a quantum spin system on a two-dimensional square lattice, will it have

[*1] Here, although we are considering the ability of prediction/simulation as one basic measurement for "understanding", it should be noted that prediction is not the *only* aspect of understanding. In fact, *explaining*, which is the other aspect of understanding seem more essential. However, quantitative arguments on explanation seems far more difficult than dealing with predictions and simulations. While the attitude of considering simulation as the measure of understanding resembles the attitude towards intelligence in the Turing test [1], it may also be criticized to be too instrumentalistic [2]. A computational complexity approach for making quantitative arguments of understanding has been discussed in [3].

a finite energy gap in the thermodynamic limit? Or given a dynamical system, will it ever get close enough to a certain configuration? These problems are known to be *undecidable*, meaning that there exists **no** finite procedure for correctly answering those questions in general [5, 6]. We will also review a more subtle hardness known as NP-hardness, for the problem of determining the ground state of a given spin-glass system, in this thesis. All of the examples above seem to have *some* connection to the actual complex behavior physicists need to tackle. More precisely, the hardness of determining the spectral gap of quantum spin systems was one of the very reasons why the Haldane gap [7] became such a controversy. It is tempting to understand the chaotic behaviors occurring in various dynamical systems as a reflection of their undecidability. The very fact of the problem of finding the ground state of a spin glass seems related to the slow relaxation of those systems.

Computational complexity, or more broadly the theory of computation, is the mathematical field which deals with these hardnesses, and quantifies them. They reveal beautiful structures behind seemingly unrelated fields, from the perspective of hardness. The theory also formalizes concepts such as (e.g., NP-) "completeness", which could be regarded as the computational complexity analogue of "universality classes" in physics, arising universally in apparently different situations.

It would be impossible to address all the hardnesses arising in physics here. While it is interesting to consider each case of the interrelations between hardness and physics, the way they arise varies very much. Even for the examples we have seen so far, there were not only cases where the hardness was a mere obstruction, but also cases where hardness played a positive role, and even cases where hardness may be used as an explanation. The main purpose of this Ph.D. thesis is to address those diverse relations between physics and hardness, or computational complexity, through our main topic, Quantum Annealing.

Quantum Annealing

One topic of hardness was obviously missing from the above examples; quantum theory. The hardness of simulations of quantum systems was originally pointed out by R. P. Feynman [8], D. Deutsch [9], and other physicists. The revolutionary idea was to utilize this hardness and use it to compute hard problems. This lead to the concept of quantum computers, which is being pursued today globally. Computational complexity theory clearly plays an important role here: distinguishing what problems are hard and what are not; or what is quantum and what is classical.

Quantum annealing (QA) was proposed from a somewhat different motivation originally, as a quantum counterpart of the physics-inspired "simulated annealing" algorithm. Importantly, adiabatic quantum computation (AQC) which was a physical protocol proposed few years after QA, and was very close to QA[*2], turned out to be equivalent to quantum computers. This means that QA and/or AQC could be analyzed from two different aspects, one as a physical device, or a physical protocol/phenomena, and one as a quantum computer. This made QA an almost ideal subject for studying the interplay of physics and computation, since most of the ways of interactions are covered to some extent. Some naive questions which may immediately arise considering this situation are: can quantum computers solve the NP-hard problems which are ubiquitously found?; do they truly exceed classical computers?; etc. Our main motivation in this Ph.D. thesis is to answer these questions, which naturally arise when looking at QA from the computational complexity point of view, as far as possible.

Furthermore, it should be worth mentioning that recent progress in quantum control

[*2] Perhaps, "a slight extention" may be a better explanation, which we will discuss thoroughly in chapter 3.

technology achieved large scale (~ 2000 qubits) implementation of QA. As we will see in the following chapters, the quantum annealing machine which has been realized so far, is a restricted version of the maximally powerful framework, and is believed not to be equivalent to quantum computers in the strongest sense. The quantum computer on the other hand, is under development. Experiments are proceeding actively, and announcements of implementation up to ~ 50 qubits seems plausible within a year time scale [10].

Structure of this thesis

As we mentioned, the main purpose of this thesis is to fully address the interrelations of physics and computational complexity through the topic of QA. The interrelation could be roughly classified into two types, one where physics affects computational complexity, and other where computational complexity affects physics. We would also like to address the hardness arising from quantumness. The three main chapters in this thesis will each deal with one type of the relations listed above.

First, we will review computational complexity theory in chapter 2. The field serves as the common language for describing hardness. Basic concepts such as completeness, reductions, and complexity classes will be defined. We will see how computational complexity can have implications for physics *at all*, by introducing the Church-Turing thesis together with its "physical" versions, and also the NP hardness assumption, which we will set as a standpoint in later chapters.

In chapter 3, we will explain both QA and AQC. As we mentioned, their relations are subtle and we will explain what we mean exactly by "QA is equivalent to quantum computers *in principle*", which is a commonly used phrase. We also define "stoquasticity" which will become one of the central concepts when we consider classical simulability of QA.

In chapter 4, we address the question "is QA really *not* efficiently simulable by classical algorithms", from the perspective of Monte Carlo methods. While the stoquastic property of the QA Hamiltonian allows us to conduct classical simulations, we show that the equilibration time for the simulations may grow exponentially. We review examples called "obstructions" where it is shown that the classical Monte Carlo algorithm requires exponential time for equilibration while QA will only need polynomial time. Then we will provide arguments and also numerical evidence on how the obstructions may be overcome by using slightly more sophisticated algorithms. We present the seemingly "most likely way of proving classical simulability", which uses the exchange Monte Carlo (EMC) method. Then we construct an example which makes this approach fail, and numerically demonstrate so. This chapter will contain a detailed explanation of the Monte Carlo techniques which we will use in the later chapters. At the same time, we will be seeing common features of such algorithms which requires stoquasticity. This chapter will serve as making clear the fact that stoquasticity indeed lies somewhere between classical and quantum, and demonstrate how difficult it is to prove either equivalence or separation. This will motivate considering the setting for chapter 6, where stoquasticity may indeed have some physically interesting property.

In chapter 5, we address the question "how does physics affect computational complexity", in a classical mechanics set up. We first review known results on physical phase transitions which occur in *problems*. The phase transitions are *spin-glass* transitions, which we will discuss its properties. While it may seem obvious that the spin-glass transition induces computational hardness, this correspondence (which we will call the RS-RSB/easy-hard correspondence) is not known to generally hold true. In this chapter, we focus on the maximum independent set (MIS) problem which is an NP-hard problem, which previous study suggested the break of the correspondence. We construct a new algorithm, which works exponentially more efficiently in some parameter region. The novel algorithm will work efficiently up until the spin-glass phase transition point, exhibiting the RS-RSB/easy-hard correspondence. This chapter will show the necessity of an adequate algorithm to explore the relations between computational complexity and physics. This chapter also serves as a guiding principle in the next chapter, when we want to construct average-case hard ensembles for the MIS problem. We use knowledge of average-case complexity introduced in chapter 2, both in this chapter and in the next.

In chapter 6, we address the question "how does computational complexity affect physics", or more precisely, the question "does computational complexity have any implications for physics". We first argue that the NP hardness assumption will lead to a physical prediction for QA, namely that it predicts the existence of an exponentially small energy gap for some Hamiltonians. This by itself is straightforwardly deduced from the NP hardness assumption, and what makes it more interesting is the possibility that the exponentially small energy gap may accompany some physical phenomena. If this is the case, a computational complexity assumption will lead to a physically novel prediction. This is totally possible, and we see this from reviewing two previously proposed physical pictures for explaining the physical backgrounds of the exponentially small energy gaps. We use Monte Carlo techniques introduced in chapter 4 for analyzing the ideal QA procedure in the adiabatic limit, for the MIS problem with unique solutions. Insights and algorithms presented in chapter 5 will be used for constructing hard-on-average MIS problem instances with unique solutions. We will find first order transitions which accompany exponentially small energy gaps. Furthermore, a novel quantum phase will be detected only by the fidelity susceptibility χ_{F}. We confirm that the novel phase is compatible with *neither* of the previously proposed physical pictures. We especially see that the novel phase is not a spin-glass phase from observing the spin-glass susceptibility for a slightly different ensemble, which the classical limit is known from studies done in chapter 5. While this chapter will leave a big open question about the novel quantum phase, it will show the possibility for a novel physical phase arising in a computational setting.

In chapter 7, we will go back through the chapters and discuss possible future directions, and finally conclude this thesis.

Chapter 2

Backgrounds from Computational Complexity Theory

We will first start with reviewing computational complexity theory. Complexity theory deals with how hard problems are, and enables us to compare problems both quantitatively and qualitatively. It was in the 1930's when the field was established with consecutive works of A. Turing, A. Church, K. Gödel, S. Kleene, and others[11, 12, 13].

One of the most important implications to physicists from this field, even in today, is the Church-Turing thesis, and perhaps its extended versions [14, 9, 15]. Their claims will be explained in the following sections, and readers should realize that its nature is actually closer to that of laws of physics, rather than mathematical statements, and thus why it is interesting for physicists.

It should be noted that the author intends to write a minimal but sufficient review in order to fully understand the main topics of the present thesis in this chapter. If a more thorough review is needed, the author recommends [16, 17, 18].

2.1 Preliminaries: Classical Complexity

We first introduce basic notions of classical computational complexity, i.e. the non-quantum part and also in the sense that we only focus on the very "classical" aspect of it such as decision problems and worst-case analysis. Other frameworks and connections to those will be addressed in later sections.

2.1.1 Classical Computation Models

The very definition of computation is the most crucial point in computational complexity theory. For instance, it is proven [19] that if one can add, subtract, multiply, divide, and round up arbitrary real numbers in unit time, extremely hard problems (PSPACE-complete problems, to be precise) become solvable in polynomial time, which is very unphysical as we will see later in this chapter. This means that if we fail to define a physically natural concept of computation, the results it produces are physically void.

Fortunately, numbers of natural and physically sound definitions of computation were proposed in the 1930s, which surprisingly, all turned out to be equivalent in a precise sense. In this section, we will only introduce the two most popular models of computation, although other models such as lambda calculus [12], μ-recursive functions [13], and their equivalences are interesting by their own means.

The fact that seemingly very different formulations of computation turn out to be equivalent suggests that those definitions are fairly robust, and today the term "computable functions" implies functions computable by any of those equivalent formulations.

The Turing Machine

Turing machines were introduced by A. Turing in 1936 [11], becoming the most famous and intuitive computation model. A Turing machine could be understood as an idealization of a situation where a person performs a calculation using a paper and a pen. We assume that the paper is divided into meshes where each mesh can have at most one letter written in it [*1]. These letters could be additionally written or erased, depending on what the person is thinking at that moment (the *inner state* of that person). Whenever the person feels that the computation is done, i.e. when the inner state is a specified *terminal state*, the protocol terminates.

All of the above intuitive protocol can be expressed mathematically, as below.

Definition 1. Turing Machine
A Turing machine is defined as a 3-Tuple $\{Q, \Gamma, \delta\}$ with each element having the following properties.
(i) Q is a finite[*2] set of *inner states*, including an *initial state* q_0 and a subset $F \subset Q$ of *terminal states.*
(ii) Γ is a finite set of *alphabets*, including the *blank symbol* $\sqcup$.
(iii) $\delta : Q \times \Gamma \to Q \times \Gamma \times \{+1, -1\}$ is a mapping.

The brilliant point of this formulation is that it distinguishes the computer itself and the input to it. Depending on what is written on the paper initially (i.e. the input) , the person can compute different instances of the same problem. This could be formulated as below.

Definition 2. Input to a Turing machine
An input to a Turing machine $M = \{Q, \Gamma, \delta\}$ is defined as a mapping $I : \mathbb{Z} \to \Gamma$ with only finite support $S \subset \mathbb{Z},\ |S| < \infty$ which has a non-blank symbol as the image. (i.e. $\forall m \in \mathbb{Z}\backslash S,\ I(m) = \sqcup$)

Definition 3. Computation done by a Turing machine
We consider the computation done by a Turing machine $M = \{Q, \Gamma, \delta\}$ for an input $I : \mathbb{Z} \to \Gamma$. The *configuration* of the Turing machine M is a 3-Tuple $\mathcal{C} = \{q, x, I'\}$, where $x \in \mathbb{Z}$ is the position, $q \in Q$ is the current inner state, and $I' : \mathbb{Z} \to \Gamma$ is the mapping for the paper at the moment.
The configuration starts from $\mathcal{C}(0) = \{q_{\text{init}}, 0, I\}$. If the configuration at time step t is $\mathcal{C}(t) = \{q, x, I'\}$, then the configuration at time step $t + 1$ would be

$$\mathcal{C}(t+1) = \{\delta_Q(q, I(x)), x + \delta_{\pm}(q, I(x)), I''\} \tag{2.1}$$

where $I'' : \mathbb{Z} \to \Gamma$ is same as I' except for the image of x, which is now $I(x) = \delta_\Gamma(q, I(x))$. The mapping δ_* with subscripts denotes the function δ with a projection to the corresponding space $*$. The computation is over whenever the inner state of the configuration at time t has a terminal state $q_{\text{end}} \in F$ as the inner state.

The function $I : \mathbb{Z} \to \Gamma$ is the mathematical expression for the paper we mentioned

[*1] It surely is unreasonable to assume that a person can write infinitely small letters.
[*2] It is also unreasonable to assume that there are infinite numbers of inner states, since then, the person could just calculate without the paper in the first place.

above, and it is often called the *memory* [*3] of the Turing machine. The assumption that the domain of I is the entire integer $\mathbb{Z}$ represents that we have infinite memory. This assumption will be later replaced by a more physically plausible one shortly.

There are mainly two different ways of interpreting the output of a Turing machine. One way is to prepare two different terminal states, e.g., $F = \{q_{\text{accept}}, q_{\text{reject}}\}$, corresponding to outputs Yes and No. Depending on which of the two states the Turing machine is at the end, one can decide whether its output is Yes or No. The other way is to simply read out what is written on the memory at the end, and interpret it as the output. The selection of the way of output depends on the formulation of the problem in interest, and we will come back to this point later in section 2.1.2.

For now, the important point is that the Turing machine is a mathematical model of a person calculating using a paper and a pencil, and it uses some amount of time and memory which can grow as large as it needs to. We should be careful that since the above definition only provides us the *procedure* of the computation, meaning that one is not guaranteed to finish in finite time necessarily. Therefore, computation defined here is similar to a mathematical mapping, but there is an important difference between them. While an image for a mapping is always defined as long as the input is in the domain, there need not be any output for certain inputs for a computation, since it could be possible that the computation runs forever on the Turing machine.

Circuits

Circuits are also intuitive models of computation, where it is no longer a model of a person with a paper and a pencil, but closer to a digital computer in some sense. Now the input is expressed as a string of 0s and 1s, which we call *bits*. Circuits are composed of *gates* which is a simple mapping of bit(s) to bit(s). We could think of a circuit as a directed graph with no directed loops [*4], with each vertex being a gate. The input is put into the *input vertices*, which have no incoming edges. The output is simply the final bit (or the bit *string*, if there are multiple bits), for the *output* vertices, which have no outgoing edges. We can either read out the entire string as the output or just focus on one bit interpreting 1 as Yes and 0 as No (or vice-versa). These correspond to the two ways of output for the Turing machine.

One is able to compute by designing the order and position of the gates. We usually allow AND gates (represented by $\wedge$), OR gates (represented by $\vee$), and NOT gates (represented by $\neg$). This set of gates is called a *universal set* of gates, which means that by combining these gates appropriately, one can build circuits for arbitrary Boolean function $F : \{0,1\}^N \to \{0,1\}^M$. We present an example of a circuit where you can input a number up to 7 (in binary) and the circuit outputs whether if the input is a prime number or not.

When we were considering Turing machines, the focus point was on whether if a Turing machine computing a problem A *exists or not*. We cannot take the same approach for the circuit model since as we explained above, *any* Boolean function has a circuit. Therefore, we naturally have the class of circuits which have a Turing machine that outputs their blueprint, given the problem size as input. Although there are classes for circuits where we solely consider the existence forgetting the construction (these circuits are called *nonuniform* circuits) and is interesting for many reasons, in this thesis we will only consider *uniform* circuits (circuits producible by Turing machines in the way mentioned above) as they suffice to explain our motivations and define quantum classes.

[*3] or sometimes the *tape* since it is one dimensional.

[*4] This is contrary to the case of real circuits where there can be loops. Also, even in this case there could be an undirected loop.

Church-Turing Thesis

The two computation models we explained so far may seem to be somewhat artificial and have arbitrariness. For instance, one can think of having multiple memories such as $I : \mathbb{Z}^k \to \Gamma$ on a Turing machine instead of just one, or other gates for a circuit. However, it turns out to be that those types of changes in the computation models actually do not affect the computation power of those models. In other words, the class of "what can be computed in *finite* time" is known to be very robust against those changes in details of the computation model [18, 17]. Furthermore, the two models (and many others) actually have the exact same power for computing, and this is called *universality* of computation. The class of functions computable in finite time using those models is called simply "computable functions". The acute reader may notice that this naming tacitly assumes that the Turing machine cannot be extended further to compute a strictly wider class of functions in finite time. That is true, and the assumption is called the **Church-Turing Thesis**. The Church-Turing Thesis could be interpreted as a mathematical definition of what the word *computable* means: computable by a Turing machine or a circuit in finite time.

Obviously, if we drastically change the computation models, their computational power may change. For example, we can assume that the t-th time step of a Turing machine takes only 2^{-t} seconds, meaning that an infinite amount of time steps could be performed in 2 seconds. This version of the Turing machine, called a Zeno machine [20], can compute problems which the original Turing machine takes an infinite amount of time to compute, for example, the *halting problem* [11]. Such models are interesting in their own rights, but the reason why we define computability not with them but with normal Turing machines is simply because of physical realizability. We can in fact, actually build a physical Turing machine with negligible differences with the mathematical model. On the other hand, the Zeno machine would have to violate the laws of physics. Namely, if we write the maximum speed of the machinery for a step of Turing machine to be v, the same machinery would have to move with speed $2^t v$ at the t-th time step, which is prohibited for $t \ll 1$ by relativity, upper-bounding any material's speed with the speed of light.

There is no known physically sound model which is able to compute classes of functions exceeding the "computable functions" defined by the Church-Turing thesis at present. We can extend the Church-Turing thesis to a stronger statement that such *hypercomputation* models do not exist in the physical universe. This is a claim which may be overturned, possibly by discoveries of new physical phenomena allowing one to physically construct a more powerful computational model than the Turing machine. Thus, from its experimental falsifiability, this is a claim closer to a physical law in nature and is called the *physical Church-Turing thesis* [21].

Although the nature of the physical Church-Turing thesis is similar to that of a physical principle, like the second law of thermodynamics, it is usually not considered in the physics community as so. The most crucial reason for this is that the physical Church-Turing thesis has never yielded any further insights other than the statement itself to physics. For instance, the second law of thermodynamics, together with the other laws, can derive many nontrivial concepts and theorems such as the thermodynamic potentials, which enable us to predict further physical phenomena, other than the increase of entropy itself. On the other hand, despite the fact of the physical Church-Turing thesis not being refuted for decades, we do not yet have nontrivial physical implications from it. Two reasons for this could be pointed out, namely that (1) simply because physicists have not been paying attention enough, and (2) because physics usually deals with limited amount of space and time, whereas computability only distinguishes finite and infinite. The first point is precisely the motivation for the works done in this thesis, which we will come back again in chapter 6. For the second point, we must take into account further limitations on the space and time needed for computation, entering the realm of computational complexity theory*5, which will be explained in next section.

2.1.2 Complexity Classes and Reductions

While we have a perfectly good computation model as the Turing machines and circuits, the "computable functions" as defined above may be not of physical interest in practice. That is because even if a problem/function is computable in finite time, it does not mean it's practically solvable, since it may take extremely long time, or need extremely large memory space. For this reason, it is natural to define a notion of "practically" possible instead of sticking to fundamental possibility. The beautiful fact which computational complexity theory revealed is that even if we focus on somewhat practical aspects of computation, we still gain fundamental knowledge about problems and their hardnesses. In the following, the class P and NP are introduced, being the most basic complexity classes of efficiently computable and efficiently verifiable.

Problems and instances

We should first define *problems* in the computational complexity framework before getting into classes of problems. The most general definition of a problem would be the following.

Definition 4. Problem and Instance, size
A *problem* is a mapping from an input $x \in \Gamma^n$ to an output. An *instance* of a problem is each of the possible inputs. We define the *size* of the instance as the number of alphabets n. The output of a problem A to an instance x is denoted as $A(x)$.

For example, MULTIPLICATION could be seen as a problem where the input is a pair of integers (encoded in binary), and the output is their product (in binary). In that case $101, 11$ is an instance of size 6 (including the comma as part of the input), with output 1111.

Usually, the time and space needed for computing increase as the instance size increases. The focus is on how they scale with the problem size, especially in the asymptotic limit. For instance, ADDITION of two n digit numbers takes at most $2n$ steps of summing one-digit numbers, and thus the scaling is $O(n)$*6. Another example could be MULTIPLICATION of two n digit numbers. The naive arithmetic way of calculating takes n^2 steps of multiplying one digit numbers and then adding of at most $n^2 + 2n$ steps, resulting in scaling of $O(n^2)$. It should be noted that it is possible that there are more clever algorithms than naive arithmetic, and these scalings may improve*7. However, at least sticking to the naive arithmetics, now we are able to qualitatively compare the hardness of ADDITION and MULTIPLICATION, and claim that the latter must be harder. Notice that the framework of defining the instance sizes and concentrating on the asymptotic scaling enables this comparison.

*5 Up until this point, the topics were about computability theory.

*6 Strictly speaking, this should actually be represented in terms of the steps of a Turing machine. However, that is an unnecessary complication, and we stick to an intuitive explanation. For now, it suffices to assume that additions and multiplications of one digit numbers take $O(1)$ step for a Turing machine, and the argument holds.

*7 In fact, there are several ways to achieve better scaling than $O(n^2)$ for MULTIPLICATION. A simple *dynamic programming* algorithm [22] runs in $O(n^{\log_2 3})$ steps, and an even better algorithm runs in $O(n^{1+\epsilon})$ steps[23]. We can thus say MULTIPLICATION is only *slightly* harder than ADDITION essentially.

Efficiently Computable : Class P

The running time of the two examples of ADDITION and MULTIPLICATION had different powers, but were both polynomial. This generalizes to the class P, but before defining P, we first define different types of problems for later use. This is because formally, the class P and NP are only defined for decision problems.

Definition 5. Decision problems and Function problems
Decision problems are problems which outputs are YES/NO. Instances with the output YES are called yes-instances.
Function problems are problems with arbitrary outputs expressed by the alphabets Γ, e.g. integers.

Since the outputs of decision problems are more limited than function problems, it may appear that they are relatively easy or simple. This is not necessarily the case, and we see the connection between the two types of problems in section 5.1.2. For now, we mostly focus on decision problems for simplicity.

Both ADDITION and MULTIPLICATION were function problems. An example of a decision problem could be PRIMES where the input is an n digit number a, and the output is YES if a is a prime, and NO otherwise. The very definition of prime number gives a naive algorithm called *sieve of Eratosthenes*, which is basically trying to divide a by all numbers up to $\sqrt{a} = O(2^{n/2})$. This algorithm requires exponentially many numbers of steps, which is rather inefficient*8. The most intuitive explanation for why Eratosthenes's sieve takes exponential time is because it is basically conducting an *exhaustive search* trying out all potential candidates for answering the question. Thus, if there indeed exists an algorithm which runs in *polynomial* steps, it should involve some deep understanding of the problem, and that is why we define the first complexity class as follows.

Definition 6. Polynomial Time: Class P
The definition of a Problem A to be in class P is that there exists a Turing machine M_A with the following property:
There exists a polynomial $p(n)$, and M_A stops within $p(n)$ steps where n is the size of the instance, with the correct output.
These problems are said to be polynomially (polynomial-time) computable.

The class P is the most simple and fundamental class of problems which could be interpreted as "efficiently" computable (also said to be *tractable*). Surely, it is possible to argue that there is a gap between polynomial time computable, and what we intuitively perceive as efficiently computable*9. However, P has several properties which make us believe that it is a meaningful and essential class: it is closed under many operations*10, it is robust against minor changes in the computation model*11, a natural definition using

*8 It is true that if we encode the input a in unary (i.e. a number a is represented by a iterations of the symbol 1) the steps only scale as the square root of the input size. Therefore, the way an instance is encoded should be carefully defined, depending on what the focus is on.

*9 For example, polynomials with huge degrees or huge coefficients may turn out to be as worse than an exponential for real-life instance sizes.

*10 This includes e.g. intersection, union, addition, oraclek with its self [18].

*11 Indeed, if a Turing machine with k tapes can compute something in time T, it can be simulated with a two-tape Turing machine with time $T \log T$ [24], and a two-tape Turing machine running for T steps can be simulated a single-tape Turing machine with time T^2. Thus it seems that powers of polynomials may change by the computation model but the boundary of polynomial/exponential first-order logic very different from the one above exists[25], and finally, most of "natural-arising" problems run in relatively small degree polynomials empirically[3].

It should be noted that the definition of P requires only the *existence* of such Turing machine M_A. Thus, even if a problem does not have a polynomial time algorithm at present, it does not rule out the possibility of that problem being in P. In general, proving a problem to be outside of class P could be extremely difficult, and the fact alone that there does not exist any polynomial algorithm yet, should not be a convincing evidence. In fact, conversely to expectations, it turned out to be that PRIMES is in P in 2002 [26].

We would also like to mention that the class if problems computable by uniform polynomial-sized circuits, exactly coincides with P.

Efficiently Verifiable: Class NP

The class P was (one of) the classes capturing the intuitive notion of efficiently computable. Another central complexity class is NP, which could be said that captures the notion of efficiently *verifiable* instead of computable.

Definition 7. Nondeterministic Polynomial Time: class NP
A decision problem A is in class NP when there exists a decision problem V such that $A(x) = \text{YES} \Leftrightarrow \exists v_{\ (|v|=\text{poly}(|x|))}\ V(x;v) = \text{YES}$, and $V(x;\cdot)$ is in P.

It should be noted that NP is defined in a different manner compared to P. Instead of looking for an efficient algorithm for solving the problem, the definition looks for an efficient way of verifying the yes-instances. The decision problem V in the definition above is the verification process that x is indeed a yes-instance of problem A.

To see this, let's see the decision problem of FACTORIZATION as an example. The input to this problem is a tuple of three integers $x := (N, a, b)$ (in binary) and the task is to decide whether there exists a nontrivial factor of N in the range $[a, b]$. This problem may seem similar to PRIMES, but is slightly different. Although PRIMES is in P, and we can decide whether a given number N is prime or composite in polynomial time, this does not necessarily mean that we can actually *find* the factors if it *is* composite. In fact, there is no known polynomial algorithm which solves FACTORIZATION*12.

Although we do not have an efficient algorithms for solving FACTORIZATION, we are able to *verify* efficiently for yes-instances, once the solution is given. We can define a problem $V_x := \text{ISFACTOR}_{N;a,b}$, which the input is an integer k, and the output is YES if k is a nontrivial factor of N in $[a, b]$, and NO otherwise. Evidently, an instance x of FACTORIZATION, is a yes-instance if and only if there exists an instance k where ISFACTOR_x outputs YES. Note that the problem ISFACTOR itself is very easy, and is in P, while finding the yes-instance may be enormously hard. The yes-instance for V_x is called a *proof.* So we can say that after all, FACTORIZATION is nothing but a problem of *finding* a yes-instance for a P problem, namely ISFACTOR. This is what the definition of NP is saying, that there is a polynomial-time proof verification algorithm for all yes-instances*13.

In short, we can say that the class NP is "the class of problems which suggested a solution, we're able to verify it in polynomial time". While it seems obvious that NP is a

remains, as long as the model is natural in some sense.

*12 Actually the cryptography method we rely on today for internet payments etc. is based on the assumption that FACTORIZATION is hard [27].

*13 The reader should not be so much bothered with the term *nondeterministic*. Historically, the class NP was defined through nondeterministic Turing machines running in polynomial time. Today, defining through a verification scheme is much more intuitive and relations with other complexity classes becomes clear. Intuitively, a nondeterministic Turing machine can miraculously *guess* the proof and just conduct the verification.

wider class than P (NP$\supseteq$P is almost trivial), it is open whether if the inclusion is strict. While it is widely believed among the majority of researchers in the field that P$\neq$NP, and numerous mathematical results are found upon the assumption that P$\neq$NP, there are few experts opposing the view [28], and resolving the problem is a central objective of computer science.

Conjecture 1. P$\neq$NP

Comparing hardness: Reductions

In the last section, we saw that P$\neq$NP remains a conjecture. If it is generally hard to prove that a problem is outside of P, what good does computational complexity provide us after all? The technique of *reductions*, and the notion of *completeness* allow us to have good reasons to believe that some problems are indeed not included in P, even though we cannot completely prove so. This also turns out to be the first supporting evidence that P$\neq$NP. Reduction enables us to compare two problems A and B in terms of hardness. We do so by considering a mapping as follows.

Definition 8. Polynomial Time mapping reduction
A decision problem A is said to be polynomial-time (mapping) reducible to another decision problem B when there exists a mapping $f : \Gamma^* \to \Gamma^*$ with the following property.
$A(x) = \text{YES} \Leftrightarrow B(f(x)) = \text{YES}$ and $x \mapsto f(x)$ is computable in polynomial time of $|x|$.

Let's assume that a mapping as defined above exists for two problems A and B. In this case, an algorithm computing problem B in polynomial time implies also an algorithm computing problem A in polynomial time. It is not so hard to see why. If one wants to compute $A(x)$, the first thing you do is to compute $f(x)$. Now that you have an instance of B, you can use the polynomial time algorithm to compute $B(f(x))$. The fact that $f(x)$ is polynomial time computable in $|x|$ means that the length of $f(x)$ is also at most polynomially long. Since a polynomial of a polynomial is also a polynomial (this is another closure property in effect), $B(f(x))$ is polynomially computable in $|x|$. By the assumption, the answer for $B(f(x))$ is equivalent to that of $A(x)$ and we obtain the desired computation. In contrast, the opposite does not hold: even if there is an algorithm which computes problem A in polynomial time, that does not necessarily mean that B is also polynomially computable. The situation could be explained as "problem B is *at least as hard as* problem A", and we write as

$$A \leq_{\text{poly}} B. \tag{2.2}$$

The "poly" in subscripts means that we are now forgetting about factors up to polynomial. Thus whenever we have a polynomial mapping reduction as defined above, it could be interpreted as problem B being *expressive* enough to *capture* the hardness of problem A. In this case, any attempt to design an algorithm solving problem B in polynomial time inevitably involves a polynomial time algorithm also for problem A, and thus it is said that problem A is (polynomially) *reduced* to problem B.

NP-complete/hard problems

By using the concept of reductions, we are able to define the class NP-complete. We first define one decision problem.

Definition 9. SAT: The satisfiability problem
Input: A boolean formula $\phi(\boldsymbol{x})$ with N varibales $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$
Output: Is there an assignment of $\{0, 1\}$ for $\boldsymbol{x}$ such that $\phi(\boldsymbol{x}) = 1$?

This problem is indeed in NP. A problem V_ϕ to output $\phi(\boldsymbol{x})$ for an input $\boldsymbol{x}$ is polynomially computable, and SAT is a problem asking whether this problem V_ϕ has a yes-instance at all.

Theorem 1. (Cook-Levin Theorem). $\forall A \in$NP, $A \leq_{\text{poly}}$SAT [29, 30]

We will not go into the details for the proof of this theorem, but intuitively, the proof shows that the definition of a problem A being in NP, could be formally expressed as a boolean function of at most polynomial length. The significance of the above theorem is that the single problem of SAT actually represents the hardness of *all* problems in NP. In other words, just finding one polynomial algorithm which solves SAT, would imply a polynomial algorithm for *all* problems in NP, and thus proving P=NP.

Soon after this theorem was found, there were an enormous amount of different problems which had the same property as SAT of capturing the hardness of the entire class NP. For instance, the restriction of SAT to only allow inputs in "conjunctive normal form with each clause containing at most 3 literals"*14, now known as 3-SAT, also sufficed for the Cook-Levin theorem. These problems are called either "NP-hard" or "NP-complete", the latter meaning that the problem itself is also in NP*15.

The concept of NP-completeness has a number of significances. First of all, the abundance of NP-complete problems in wide range of fields was certainly unexpected, and showed that there is a universal structure among problems in various fields which human struggles. Secondly, the fact that all of the NP-complete problems in different fields have resisted all the attempts to find an efficient algorithm, provides us an empirical evidence (which of course is not a proof) that P≠NP. Thirdly, and most important for the present thesis, is that it enables us to guarantee hardness of a problem without proving so. More precisely, if NP-hardness is shown for a specific problem, we can safely assume that the problem is not in P, as long as conjecture 1 holds.

Physical Church-Turing Thesis

We have already mentioned the Church-Turing thesis, and also the attempt to extending its claim to the physical world. The statement deals with computability, which is intrinsically a notion with unbounded computational time and memory. Then in the previous section, the classes P and NP were introduced (within computable functions), characterizing the practically computable and practically verifiable. It is a natural attempt to ask if there is a natural analogue of the Church-Turing thesis considering the efficiently computable classes instead of the class of computable.

Indeed, there are several attempts. It is tempting to think that any problems computable efficiently in the physical world, is also efficiently computable by a Turing machine, thus is in P. A very similar statement is known to be the "Strong (complexity-theoretic) Church-Turing thesis"[31]*16, which ultimately seems to be wrong. The reason is because

*14 For example, a valid input boolean function would look like $\phi(\boldsymbol{x}) = (x_1 \vee \neg x_2 \vee \neg x_3) \wedge (x_2 \vee x_3 \vee \neg x_4) \wedge (\neg x_1 \vee x_3 \vee x_4)$.

*15 So "NP-hard" means the problem is at least as *hard* as the entire class NP, and "NP-complete" means that the problem is *completely* hard as it could be with NP.

*16 The actual statement claims the efficiently computable in the physical world is in BPP instead of P. We will define the class BPP later, and for now, it suffices to know that BPP is almost the same as P but to allow probabilistic operations, and is widely believed to be actually equal to P.[32]

we believe some *quantum* processes in nature are hard to simulate efficiently by a Turing machine, since Turing machines have intrinsically classical nature. This leads to the notion of "*Quantum* strong (complexity-theoretic) Church-Turing thesis"[33] where it states the *quantum* Turing machine should be able to efficiently compute any efficient computation in the physical world.

Thesis 1. The Strong Quantum Church-Turing Thesis
Any physically efficient computational phenomena, in the sense that the physical world allows us to build a physical computer from that phenomena, should be efficiently simulable by a quantum Turing machine.

So far, we have no physical phenomena that violate this thesis. We will discuss these quantum computations in section 2.3.

Another perspective pointed out by S. Aaronson is that, while we may be able to gain increasingly strong computational power by mobilizing new physics*17, it is most likely that NP-complete problems will remain intractable, due to somewhat philosophical reasons[34]. We will avoid discussing all of his points in detail here. One point worth mentioning, which Gödel also seemed to notice [35] is that being able to solve NP-complete problems using some physical device will mean an *automated* NP solver, which could be seen as a metaphysical consequence that we will be able to compress and understand any structured or unstructured systems in general. Thus the **"NP hardness assumption"** is stated, claiming that *no* physical system will be able to compute NP-hard problems efficiently.

Assumption 1. NP hardness assumption
There is no physical phenomenon which enables us to compute NP-hard problems in polynomial time.

Our position in this thesis is that we accept the quantum strong Church-Turing thesis, and see what physical insights we could gain from it applying it to a situation where we try to solve an NP-hard problem. We also believe that even quantum computers cannot solve NP-hard problems efficiently*18, and thus it could be said that we also adopt the NP hardness assumption. As we pointed out in the section of the Church-Turing thesis, although these statements regarding the connection of computational complexity and physics have been proposed (mainly from the computer science side), we still do not have a satisfactory application to physics. All studies in this thesis are motivated by this connection, and we will explain them at the beginning of each corresponding chapter.

2.2 Average Case Complexity and Statistical Physics

All the basic concepts introduced in the previous chapter (i.e. the classes P, NP, reductions and completeness), were based on *worst case analysis.* Recall that the definition of P was that the computation time was *upper bounded* by some polynomial of the instance size. This obviously is a definition regarding the *worst* instances.

While the worst case analysis is relatively easier to handle mathematically and has a nice structure, it may not always reflect the actual hardness of problems in the real world. This is because the worst instances may have very rare appearances in reality, with typical instances much easier.

There are mainly two different ways of coping with this matter. One is to formulate a

*17 So far, we only have quantum mechanics as the new physics but we could also think of any future physical theories which provide us with new computational power.
*18 This could be simply stated as $\mathsf{NP}\not\subseteq\mathsf{BQP}$. We define the class BQP in section 2.3.

similar framework for average-case analysis. However, it turns out that there is a major obstacle in this direction, and considerably less is known about the structures of the classes in this framework. This could be seen as a rather top-down approach for average-case. The other way is a bottom-up approach, where we set a simple probability distribution among the instances for a specific problem, and analyze their hardness according to the probability distribution. The framework for statistical physics actually suits perfectly for this approach.

In the next two subsections, we will briefly review the two approaches, and see the connections of average-case complexity and statistical physics.

This problem was realized soon after the worst case analysis method succeeded with the birth of the notion of NP-completeness. Levin was the first to tackle the problem and define complexity classes for the average-case framework [36].

2.2.1 Average-case Analysis in Computational Complexity

As in the case of worst-case analysis, we first define the class of average-case efficiently computable. L. Levin was the first to tackle the problem and define complexity classes for the average-case framework [36], soon after his findings of NP-completeness for the worst-case analysis. We will see this formulation in the following. In the following, we denote the time it takes for an algorithm R to solve an instance x of a problem A as $\mathrm{Time}[R(x)]$.

Formalism

It is interesting that even the attempt to *define* average-case tractability is nontrivial. It may appear that simply taking the average (mean) value of the computation time,

$$\mathbb{E}_n[R(A)] := \frac{1}{2^n} \sum_{|x|=n} \mathrm{Time}[R(x)], \tag{2.3}$$

would be a good measure indicating the typical complexity of problem A. However, this naive formulation fails to capture the robust notion we expect as in P for the worst-case.

Let us see why this is, by defining an "average-P" class for all the problems having an algorithm R with polynomially growing $\mathbb{E}_n[R(A)]$. We then think of an algorithm R' which is slower with a squaring factor than R for all the instances, i.e. $\forall x, \mathrm{Time}[R'(x)] = \mathrm{Time}[R(x)]^2$. Since the difference between these two algorithms is only polynomial, it is naturally expected that if R is an "average-P" algorithm, then R' is so as well, and vice-versa. However, this does not hold in general. The following example illustrates this. Let's say that R computes all 2^n instances with size n in polynomial time $P(n)$, except for one instance where R needs 2^n time. In this case, the mean time calculated by eq.(2.3) is polynomial for R, but is exponential for R'.

This problem essentially comes from the difficulty of evaluating instances with *extremely low appearances but with extremely long computation time*, especially when the rarity and the computation time are quantitatively comparable[*19]. In order to handle this, we define a problem class of typically tractable as follows.

[*19] The same type of difficulty arises in different fields. The St.Petersburg's paradox[37] in economics could be seen as a basic model where this difficulty of having events with very low probability and very high profit plays a central role. We should also remember the practical problem of risk management for disasters with immense damage but low occurrences face the same difficulty.

Definition 10. Typically tractable: Class Avg-P
A problem A is in Avg-P when the following holds.

$$\exists\epsilon > 0,\ \exists P(n) : \text{polynomial} \ \ \text{s.t.}\ \frac{1}{2^n}\sum_{|x|=n}\text{Time}[R(x)]^{\epsilon} = O(P(n)) \tag{2.4}$$

The ϵ in the definition above plays the role of "rounding off" the exponentially small probability of facing a hard instance. For instance, even if we only have an algorithm like R' for a problem A, A is in Avg-P since we can set $\epsilon = 1/2$. This definition stands in the position that an exponentially hard instance should be ignored as long as it only has a comparably small probability, which sure is not a *typical* instance*20 A definition which makes more clear of this viewpoint is introduced in [36], where the probability of an instance having a running time longer than t is bounded by $P(n)/t^{\epsilon}$ with some $\epsilon > 0$. This definition turns out to be equivalent with Avg-P.

So far, we have implicitly assumed that the probability of each instance with size n appearing has the same value $1/2^n$, but obviously this is not always the case. Thus in average-case analysis, a problem A together with its distribution $\mathcal{D}$ over the instances should be specified, and the tuple $(A, \mathcal{D})$ of these two are called *distributional problems*. We can naturally extend the definition of Avg-P to distributional problems by simply taking the expectation values according to the distribution $\mathcal{D}$ instead of the uniform distribution.

Current state

As average-case tractable problems were defined, we can ask if an NP-complete problem is indeed hard on average under some distribution. Of course, proving a distributional problem to be outside of Avg-P is even harder than proving a problem to be outside of P ($(A, \mathcal{D}) \notin$Avg-P implies $A \notin$P). Therefore, a natural strategy would be to find an analogous class of NP-complete for distributional problems, and showing that a problem to be in that class would imply intractable on average.

The class dist-NP-complete is defined in such a way*21. However, compared to the worst-case NP-complete problems which abundant problems are known, there are only few known dist-NP-complete problems. This is why relatively little amount of research pushes further this direction so far. A natural distributional problem (such as k-SAT for specific natural distributions) found to be dist-NP-complete would be a major breakthrough.

2.2.2 Average-case Analysis in Statistical Physics

As we saw in the previous section, a top-down approach such as dist-NP-completeness seems to be hard to apply for natural arising problems. We review a bottom-up approach from statistical physics which may be a good alternative.

Statistical mechanics for spin glasses, and random systems in general, resembles the situation of a distributional problem. The Hamiltonian of spin glass has random variables, and the system is set only after fixing the random variables. This is called *quenched* disorder, and the systems after fixing the quenched disorder are called *samples*. We can define the probability for each sample appearing. This corresponds to the distribution in the average-case complexity, and samples correspond to the instances.

*20 Thus, "*typical*-case complexity" seems to be a more suited term than "*average*-case complexity".
*21 In order to do that, reductions among distributional problems should be defined first, which we will not get into the details.

As an example, we can think of a system with Ising degrees of freedom $\sigma_i \in \{\pm 1\}$ with the Hamiltonian below.

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j \tag{2.5}$$

Here, J_{ij} are the random variables, and one can think of various different distributions for J_{ij}. For example, if each J_{ij} is picked from a Gaussian distribution*[22], this becomes the Sherrington-Kirkpatrick (SK) model, the most basic model for spin glasses[38]. Another distribution could be $\mathrm{Prob}(J) = \frac{1}{3}\delta(J+J_0) + \frac{1}{3}\delta(J) + \frac{1}{3}\delta(J-J_0)$, which means each pair of Ising spins have an interaction with the same strength but varying in signs.

Importantly, when we put the probability for samples aside, the following problem is known to be NP-complete [39].

Definition 11. IsingSpinGlass
Input: $N(N-1)/2$ numbers $\{J_{ij}\}_{1\leq i<j\leq N}$ all from $\{0,+1,-1\}$ and an integer k
Output: Is the ground state energy of $H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j$ lower than k ?

Classical statistical mechanics predicts that at temperature T, a system in its equilibrium state can be modeled as having a probability distribution over all possible configurations with $\mathrm{Prob}[\boldsymbol{\sigma}] \propto e^{-\beta E(\boldsymbol{\sigma})}$ (β is the inverse temperature $1/k_B T$ with k_B being Boltzmann's constant. Through out this paper we will set $k_B = 1$.). Thus, the above problem being NP-complete means that equilibrating at low enough temperature is as hard as solving an NP-complete problem. If we follow the NP hardness assumption, this means that (at least for some samples), the time for a system with the above Hamiltonian will be exponentially long.

We also note that since NP-completeness is a notion in the worst-case analysis, we can always generalize the problem and still have NP-completeness. For instance, an IsingSpinGlass problem with local fields with the Hamiltonian

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i, \tag{2.6}$$

will also be NP-complete.

As we state above, we can think of the varying interactions J_{ij} as *quenched* disorders, meaning that they are fixed to some value due to a random process through forming. Thus from a physicist's view point, we are interested in the *typical* behavior of such random system. A powerful tool found in statistical physics is the *replica method*, which deals systems with *two* different types of random variables. We will have a brief review on this issue in chapter 5. For now, we will mention that the replica method reveals phase transitions called *replica symmetry breaking* (RSB) in the SK model and other similar models. Since this approach enables us to calculate various quantities, it is more practical than the framework of average-case complexity. We will seek for connections in these two frameworks in chapter 5.

2.3 Quantum Computation

So far, the computation models we have discussed were all classical. It may seem peculiar to state that a computation model (and not a physical model) is quantum/classical. However, when we face that quantum theory inevitably destroys our classical worldview

*[22] The parameters are determined so that the average energy becomes extensive.

and denies local hidden variable theories, we notice that Turing machines and circuits were well within the worldview of classical reality, and a whole new field of quantum computation opens up.

2.3.1 Quantum Computation Models

As we explained the quantum strong Church-Turing thesis in §1, it was recognized in the early 1980's that some physical phenomena are *not* polynomially simulable by a classical Turing machine, pointed out by R. Feynman, D. Deutsch and others [8, 9]. While the quantum version of a Turing machine played a central role in the early days, the quantum circuit model is easier to handle and understand. In this section, we will introduce the quantum circuit model for defining quantum complexity classes.

The quantum circuit model is a natural extension of the circuit model introduced before, to a quantum setting. Remember the classical circuit model had an input of *bits* $\{0,1\}^N$ and transformed them accordingly to the gates, always having a local quantity for each vertex. Since quantum mechanics allows us to have superpositions of states, we can also think of superposition of bit strings, and we call these *qubits* short for quantum bits. These qubits are realizable by using spin-$\frac{1}{2}$ particles or other two-state quantum systems. We write the two states a qubit can take to be $|0\rangle$ and $|1\rangle$ similarly to bits, with the vector form

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \ |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}. \tag{2.7}$$

This basis is called the computational basis, or the z-basis, interpreting it as a spin. We allow the superposition of these two states, and write $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, and call them the x-basis for later convenience. Thus we have a 2^n dimensional Hilbert space for an n qubit circuit. The difference with the classical case is that the number of qubits always remain the same in the quantum case. This is because we only allow unitary transformations among the qubits, which will not change the dimension of the Hilbert space.

The quantum analogue for classical gates is unitary operators which act on at most three operators[*23]. For instance, a gate operating on the ith and $i+1$th qubit could be written as $\hat{U} = \hat{1}_1 \otimes \hat{1}_2 \otimes \cdots \hat{1}_{i-1} \otimes \hat{U}_{i,i+1} \otimes \hat{1}_{i+2} \cdots \otimes \hat{1}_n$, where each subscript denotes which qubit the operator is acting on, and $\otimes$ is the tensor product. As in the classical case, we need not think of *all* of the possible unitary operators, but a finite set suffices to achieve an arbitrary desired unitary operator to be achieved approximately, called a *universal* set of gates[*24]. Namely, the set of Toffoli gates and Hadamard gates is a universal set. They can be expressed as

$$\hat{U}_{\text{Toffoli}} := \sum_{a,b,c\in\{0,1\}} |a,b,c\oplus ab\rangle\langle a,b,c|, \qquad \hat{U}_{\text{Hadamard}} := \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}. \tag{2.8}$$

If we write down the Toffoli gate explicitly like the Hadamard gate, it will be an 8×8 matrix with all the diagonal elements 1 except for the last two, where they will have elements 1 in the "swapped" position. All the other elements are 0. After all, we can think of the computation done by a quantum circuit as a sequence of unitary operators

[*23] It is possible to reduce this number to two. The situation is similar as in classical circuits which in principle we only needed one type of gate NAND. For now, we choose three for simpler explanation.

[*24] There is a difference with the classical case where it is only the universal set in the quantum case only achieves *approximate* universality. However, we can show that this approximation can be made exponentially accurate, and does not become problematic [18].

acting on a tensor product state $|000\ldots0\rangle := |0\rangle^{\otimes n}$ as the initial state[*25]. Thus, the output state could be written as

$$\hat{U}_{p(n)}\hat{U}_{p(n)-1}\cdots\hat{U}_1|000\ldots0\rangle \tag{2.9}$$

where each $\hat{U}_l$ is either a Hadamard gate acting on one qubit, or a Toffoli gate acting on arbitrary three qubits.

The final output of a quantum circuit is the result of a projective measurement on one or more qubit(s). The basis which the measurement is done is usually the computational basis. We provide a simple example of a quantum circuit below. Note that in order to know what this quantum circuit computes, one will generally need to compute an exponentially large matrix (in this case an $2^3 = 8$ by 8 matrix).

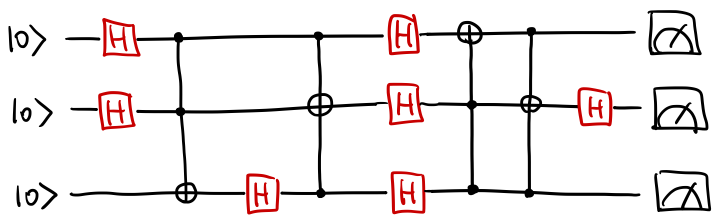

Fig. 2.1: An example of a quantum circuit with 3 qubits. The red boxes with an H symbolizes the Haramard gate, and the $\oplus$ connected with dots represent the Toffoli gate.

2.3.2 Quantum Complexity Classes

Now that we have a quantum circuit, we are able to define the analogue of P for the quantum case. However, there is a little bit of caveat for the new complexity class, which we will see by first defining the *probabilistic* analogue of P, as follows.

Definition 12. Bounded-error Probabilistic Polynomial Time: class BPP
The definition of a decision problem A to be in class BPP is that there exists a *probabilistic* Turing machine[a] M_A with the following property:
There exists a polynomial $p(n)$, and M_A stops within $p(n)$ steps where n is the size of the instance. The probability of having the correct output must be $\geq 2/3$.

[a] We will not go into the details of a probabilistic Turing machine. It will be enough to know that it is a Turing machine which is also allowed to move/write probabilistically.

While the value $2/3$ itself is not so important, the fact that the above definition allows *some* error is crucial. If we replace "$\geq 2/3$" with "$= 1$" in the above definition, tolerating no error at all, the class becomes trivially equivalent to P. Thus, the most natural way to have a complexity class with randomness is the above BPP[*26], allowing some small probability of error. Note that a different value instead of $2/3$ will end up in an equivalent class as long as the value is a constant between $(1/2, 1)$. We can always repeat the randomized computation for several times, and by taking the majority of the outputs,

[*25] We can assume the initial state to be $|000\ldots0\rangle$ without loss of generality.

[*26] There is another approach which instead of having a strict polynomial running time limit and an error-tolerate class, is strict in the sense no error is allowed but now the running time can be unbounded but with a polynomial average. This class is called ZPP (standing for Zero-error Probabilistic Polynomial-time), and it is relatively easy to see that ZPP$\subset$BPP [40].

the probability of getting the answer wrong will decrease exponentially [40][*27]. In a way, the class BPP could be seen as the natural extension of P to probabilistic algorithms, since it has an extremely low probability of failing[*28]. This way of *bounding* the error probability is standard in computational complexity theory, and the resulting complexity class becomes natural[*29].

Now that we have the definition of BPP, it is very easy to define the *quantum* analogue, BQP.

Definition 13. Bounded-error Quantum Polynomial Time: class BQP
The definition of a decision problem A to be in class BQP is that there exists a *quantum* circuit C_A with the following property:
There exists polynomials $p(n), q(n)$, and a Turing machine M_A^C which outputs the blueprint of C_A given the "size of the instance" in unary as an input. M_A^C stops within $p(n)$ steps, and C_A has size $q(n)$, where n is the size of the instance. The probability of measuring the correct output from C_A must be $\geq 2/3$.

The details regarding the Turing machine is simply the uniformity condition we have discussed for circuits in general. The important part is that we have a polynomial-sized quantum circuit, and an error-tolerance condition same as BPP. Thus, it is clear that BQP is the natural extension of BPP to the quantum case.

While it is believed that BQP is strictly larger than P or BPP, it is far from being proven. For instance, FACTORIZATION is in BQP, while no BPP algorithm is found so far. Although it is too much to hope for a rigorous proof that BPP$\subsetneq$BQP, we will stand the position that BQP is in fact stronger than BPP due to the large number of known results which are exponentially faster than the state-of-the-art classical counterpart [43, 44, 45, 46]. These exponentially more efficient quantum algorithms will be an example of *quantum advantage*[*30] in computation.

Importantly, although BQP is believed to strictly contain the classes P or BPP, it is also believed *not* to contain NP. While any separation results in complexity theory are difficult, and the situation is same for NP$\not\subseteq$BQP, there are results which strongly suggest so[*31] [49]. This is also a statement which needs to be true if we really believe in the NP hardness assumption. We show a picture of the (conjectured/strongly believed) relations between the complexity classes we have introduced so far in Fig. 2.2. Furthermore, Fig. 2.3 shows the relations of complexity classes which will be introduced in the next chapter.

---

*27 This will not be the case if "$\geq 2/3$" is replaced with "$> 1/2$" in the above definition, since the actual probability may be exponentially close to 1/2. In this case, we will need an exponential repetition for the majority vote to become reliable. This class is called PP for probabilistic polynomial, which is a way larger class even compared to NP.

*28 It sure seems meaningless to care about the exponentially small probability of failing, when it is smaller than the probability of say, an asteroid hitting the computer.

*29 The class BPP is natural enough that we have a mathematical conjecture that P=BPP [41, 32], very much intuitive for physicists since we model many ultimately deterministic processes as random events. P=BPP essentially implies that there exist some deterministic processes which the outcome is indistinguishable from that of a truly probabilistic process, as long as the distinguisher is polynomially bounded in computational power [42].

*30 The term *quantum supremacy* is often used in the literature [47]. However, there are moral problems regarding the term from historical viewpoint [48], and in this thesis, we will refer to the concept as quantum advantage instead.

*31 Both results of NP$\not\subseteq$BQP and P=BPP can be shown through *random oracle arguments*. While almost all results shown by a random oracle argument turns out to be correct, there is an exception, and thus we are bound to say only "strongly suggest" regarding results from it.

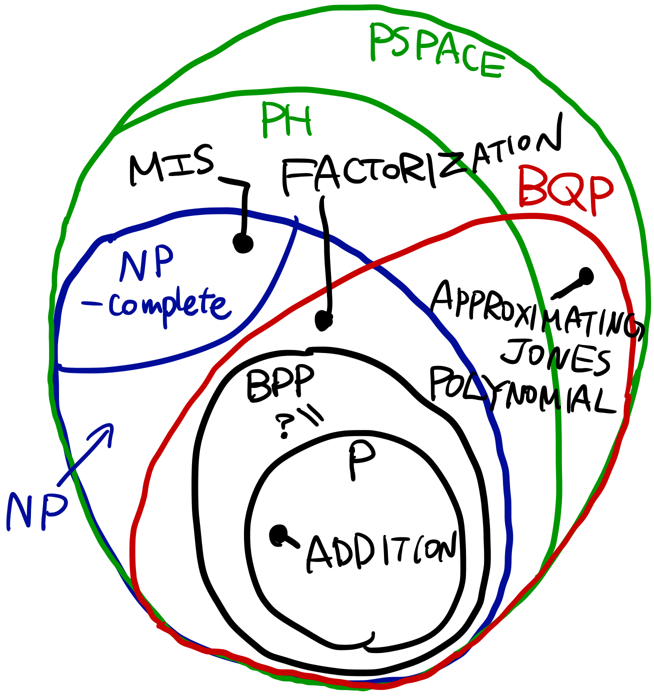

Fig. 2.2: A Venn diagram of the conjectured relations between the complexity classes. The MIS problem will be defined in chapter 5. Jones polynomials are topological invariants of knots, which somehow the approximation for a specific value turns out to be BQP-complete [50]. It should be noted that even the most separated classes of P and PSPACE in this diagram are actually not rigorously *proven* to be different so far. This indicates how difficult it is to prove separation of classes in the computational complexity field.

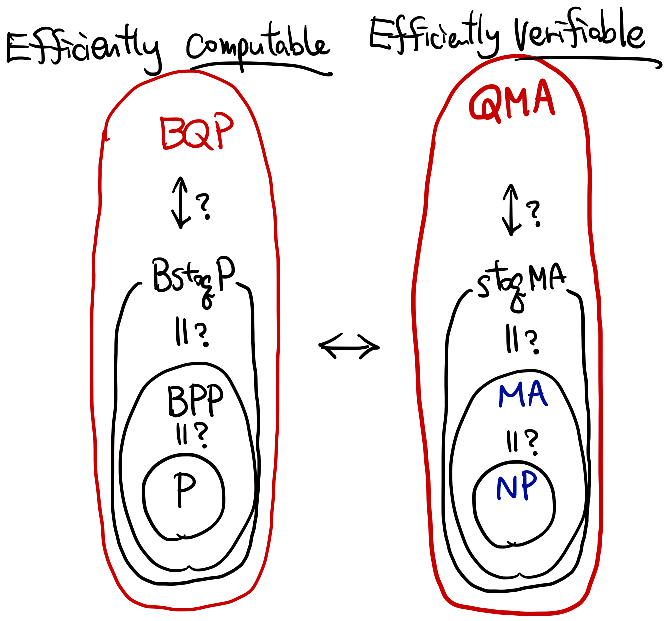

Fig. 2.3: A schematic diagram of the relations of complexity classes which will be mentioned in section 3.2.2.

Chapter 3

Quantum Annealing

Quantum Annealing (QA) has an interesting history, which started from a statistical mechanics perspective [51] and then being connected to a computationally sophisticated model [52], eventually leading to a large-scale implementation by a Canadian start-up company [53].

We will first introduce the computationally sophisticated *adiabatic quantum computation* (AQC) in the next section and then turn to QA in the following section. The term QA has been used in several different ways, and we avoid confusion by defining what we mean precisely by QA later in this chapter.

3.1 Adiabatic Quantum Computation

AQC was originally introduced by E. Farhi et al. [52], utilizing the quantum adiabatic theorem for computational use. We will first review the theorem.

3.1.1 The Quantum Adiabatic Theorem

Let us consider the quantum adiabatic theorem for systems with finite degrees of freedom and discrete state [54]. The quantum system is described by a state vector $|\psi(t)\rangle$, and the time evolution obeys the Schrödinger equation

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

We set the units so that the Planck constant $\hbar = 1$.

Theorem 2. (Quantum Adiabatic Theorem). Consider the situation of smoothly changing the Hamiltonian of a system as $\hat{H}(t)$. Suppose that initially at $t = 0$, the system is in the nth energy eigenstate $|\Psi_n\rangle$ and that the corresponding energy eigenvalue E_n is nondegenerate. Also, assume that $\hat{H}(t)$ is nondegenerate throughout the process. In this case, if the change of $\hat{H}(t)$ is infinitely slow (i.e. *adiabatic*), the state vector at time t is the instantaneous nth energy eigenstate $|\Psi_n(t)\rangle$.

If we think of the ground state ($n = 0$), the theorem suggests that under adiabatic changes of the Hamiltonian, the system will always remain in the instantaneous ground state. While the above original theorem only states about the slow limit, in reality that is not achievable, and we always have a finite time limit t_{fin}. In this case, a condition for the change in Hamiltonian to be effectively adiabatic becomes [54]

$$\frac{\max\left|\langle\Psi_1(t)|\dot{\hat{H}}|\Psi_0(t)\rangle\right|}{\min(E_1(t)-E_0(t))^2} = \epsilon \Rightarrow |\langle\Psi_0(t)|\psi(T)|^2 \geq 1-\epsilon^2, \tag{3.2}$$

where the maximums/minimums are over $t \in [0, t_{\text{fin}}]$. If we restrict all coefficients in the Hamiltonian is polynomially bounded at all time, the numerator is at most of polynomial scale. This means that as long as the energy gap $\Delta E(t) := E_1(t) - E_0(t)$ is lower bounded by some inverse polynomial $\frac{1}{p(N)}$ throughout the process, the time needed to realize the adiabatic transition is upper bounded by a polynomial $p(N)^2$. This is the main idea of the adiabatic quantum computation.

3.1.2 Formalism and Equivalence

Equipped with the quantum adiabatic theorem, we are able to formulate AQC. While it is instructive to see AQC as a general protocol/algorithm or framework, it is also possible to consider it as a computation model. Here, we will define AQC as an algorithm being able to handle a broad range of problems.

Definition 14. Adiabatic Quantum Computation

1. Prepare a Hamiltonian $\hat{H}(t)$ which has the following three properties.
   - The initial Hamiltonian $\hat{H}(t = 0)$ has an easily preparable ground state.
   - The final Hamiltonian $\hat{H}(t = t_{\text{fin}})$ has a ground state corresponding to the solution of a problem in interest.
   - The energy gap $\Delta E(t)$ is finite for $\forall t \in [0, t_{\text{fin}}]$.
2. Prepare the system to be in its ground state at $t = 0$, i.e. $|\psi(0)\rangle = |0(0)\rangle$.
3. Let the Hamiltonian change accordingly to $\hat{H}(t)$, and the state evolve.
4. At time t_{fin}, the state vector of the system is close to $|0(t_{\text{fin}})\rangle$.

The above formalism of AQC leaves a great deal of freedom to be further decided. For instance, the entire construction of the Hamiltonian is left out, and we discuss this point in the next section, relating it to QA. It is also interesting to ask what class of problems have solutions expressible by a state vector.

For now, we focus on the aspect of computation time. The above definition only ruled out cases where $\Delta E_{\min} = 0$, where we are unable to follow the adiabatic path in finite time. We can further restrict the energy gap to be at most polynomially small $\Delta E_{\min} \sim O(p(N)^{-1})$, guaranteeing polynomial time ($O(p(N)^2)$) computation. Here, N is the size of the problem, which can be also thought as the number of qubits. This enables us to consider AQC as a polynomial protocol to compute problems. Now the main issue is how we design the adiabatic Hamiltonian $\hat{H}(t)$. QA is a straightforward and general construction of the adiabatic Hamiltonian, which we will discuss in the next section.

AQC may appear as a method which just happens to use the adiabatic theorem for computation, and it may even seem strange to focus on it. One of the great reason for studying AQC is because of the following theorem [55].

Theorem 3. (Equivalence of polynomial AQC and BQP). The class of problems computable by an AQC procedure of polynomially small energy gap is equivalent to BQP.

Since this theorem could be interpreted as "$\mathsf{polyAQC}$"=BQP, its proof has two steps, namely showing that polynomial AQC is efficiently simulable[*1] by a polynomial-sized quantum circuit, and then proving that polynomial-sized quantum circuits are efficiently

[*1] The word simulable is an adjective of something being able to be simulated. We can find both "simulatable" and "simulable" in literature, with the former appearing mainly in computer science literature and the latter appeing in physics literature. In this thesis we will use simulable, since it seems grammatically more correct regarding the word origin.

simulable by a polynomial AQC. The former *must* hold if we take the strong quantum Church-Turing thesis seriously, and indeed was proved several years before the latter[56, 57]. The latter is more nontrivial, where the basic idea is to construct a Hamiltonian, which the ground state is the entire *history* of the quantum circuit starting from the input qubits to the output. D. Aharonov et al. [55] constructed such a Hamiltonian of $N+p(N)$ qubits using terms as $-\hat{U}_l \otimes \hat{1}_1 \otimes \ldots \otimes \hat{1}_{l-1} |110\rangle\langle 100| \hat{1}_{l+3} \otimes \ldots \otimes \hat{1}_{p(N)}$ which lowers the energy for configurations corresponding to a *valid* history obeying the unitary operation $\hat{U}_l$ as written in eq.(2.9). Most of the difficulty of the proof comes from showing that the Hamiltonian constructed in this way always has a polynomially lower bounded energy gap. Notice that the above term acts on at most 3+3 qubits at once. This means that although the quantum circuit model has at most 3-body interactions, in order to simulate the circuit model AQC requires 6-body interactions.

The implication of the equivalence between the two quantum computation models is considerable. The theorem tells us that whether or not a problem could be computed efficiently using quantum circuits, could be analyzed through investigating the minimum energy gap of Hamiltonians. This may cause temptation to call AQC as "a kind of quantum computer", but there is a subtlety. Circuits and AQC, both of the quantum computation models we have introduced, assume that the physical state evolves through a unitary dynamics. This is too much of idealization since in reality there are noises, and the temperature is finite. For quantum circuits, *fault-tolerant* computation has been investigated in depth[58, 59], and there are known ways to perform quantum computation as long as the noise is reduced to a constant level. On the other hand, there are no known sophisticated ways of doing the same thing for AQC so far, although several ways are currently being studied [60][*2]. In short, although we know theoretically how to build a quantum circuit robust to realistic noise, we do not yet know how to build an AQC machine with the same robustness[*3]. We would like to, however, point out that the equivalence itself is highly nontrivial, and *in fact* we can see exponentially small energy gaps for AQC corresponding to hard problems (e.g. NP-hard problems) *as if* it is preventing the NP-hard problem from being solved, violating the NP hardness assumption. Even without a complete fault-tolerant scheme, there *needs* to be an exponentially small energy gap for an AQC Hamiltonian if it is solving a hard problem (as long as we believe that NP$\not\subseteq$BQP). The reason for this comes from not the universality of AQC but rather the ability of BQP to simulate polynomial AQC. If indeed, there were at most polynomially small energy gaps for AQC of NP-hard problems, even though we might not be able to directly solve the problems with AQC due to the noise, we still can solve the problems through a fault-tolerant quantum circuit simulating AQC. This would be against our belief, and thus there *should* be an exponentially small energy gap for AQC dealing with NP-hard problems. This will become the core idea for our study in chapter 6.

3.2 Quantum Annealing

While the framework of AQC was broad allowing virtually any Hamiltonian, it would be more convenient if we have a general framework which we could automatically use for general problems. Quantum Annealing (QA) is so far the most general strategy within AQC, inspired from statistical mechanics, and enables us to implement any NP problems to an AQC framework. In this section, we will review QA and its connections to statistical

[*2] There are also claims that AQC is inherently robust against noise and temperature, but that seems to be wrong so far [61].

[*3] It is interesting that we have the opposite situation for building an actual annealer so far (disregarding robustness).

mechanics.

3.2.1 Formalism and Physical Perspective

QA was first introduced by T. Kadowaki and H. Nishimori [51] in the context of a quantum analogue for simulated annealing (SA), which is another statistical mechanics inspired algorithm for optimization. We should first briefly review SA before introducing QA.

Optimization Problems

Both of SA and QA could be seen as algorithms for solving combinatorial*4 optimization problems. A combinatorial optimization problem is a type of function problem in the form as below.

> **Definition 15.** Optimization problem
> Optimization problems are problems which given a finite set Ω of all possible *configurations* and a *cost function* $f : \Omega \to \mathbb{R}$ as inputs, the task is to output a configuration $x \in \Omega$ which *minimizes* $f(x)$.

The problem would be trivial and certainly in P*5 if the input of Ω is written as the list of all the elements, and the cost function $f(x)$ as a list of outputs for all the arguments. In this thesis, we will focus on the most common setting where x is an N-dimensional vector (thus the size of Ω is exponential in N), and we have a polynomial time protocol of calculating $f(x)$. Optimization problems with this condition are called NP-optimization (NPO) problems, appearing in a wide range of fields just like NP problems. *6

We will discuss further on how optimization problems and NP-completeness*7 relate to our study in chapter 5.

Simulated Annealing

According to classical statistical mechanics, physical quantities of a macroscopic system at equilibrium can be calculated by the expectation value over the Boltzmann distribution*8

$$P_{\text{eq}}[\boldsymbol{\sigma}] = \frac{1}{Z} e^{-\beta E(\boldsymbol{\sigma})}, \quad Z := \sum_{\boldsymbol{\sigma}} e^{-\beta E(\boldsymbol{\sigma})}, \tag{3.3}$$

where $\boldsymbol{\sigma}$ denotes the microscopic configuration of the system and $E(\boldsymbol{\sigma})$ the corresponding energy, and β is the inverse temperature as introduced in chapter 2.

It was realized by N. Metropolis and later generalized by W. Hastings [64, 65] that Markov chain Monte Carlo algorithms enable us to sample from the Boltzmann distribution. Whether if the sampling could be done efficiently or not depends on the system and parameters under consideration, and is itself an interesting problem. We will introduce the basics of Monte Carlo algorithms in section 4.1.1. Here, it suffices to know that we

---

*4 There are attempts to widen the scope of SA and QA for *continuous* optimization problems (e.g., [62]), but we will only consider discrete combinatorial cases in this paper.

*5 More precisely, it would be in FP, the analogue of P for function problems.

*6 Needless to say, in practice there are many interesting optimization problems of other types as well. e.g. in crystal structure prediction [63] even calculating $f(x)$ seems hard.

*7 In fact, the class NPO does not straight forwardly go into NP, since it is hard to verify that a certain configuration x is indeed the *optimal*. Thus, it is in the class Σ_2^{P} which we will not explain, but will see relations later.

*8 This does not necessarily mean that the actual historgram of microscopic configurations over time matches the Boltzmann distribution.

have a (classical probabilistic) algorithm to sample the Boltzmann distribution, and the time necessary for accurate sampling is characterized by the “relaxation time”.

Now, an important observation is that if we take the *low-temperature* limit i.e. $\beta \to \infty$, the Boltzmann distribution converges to the uniform distribution among all ground state configurations. This means if we have a system with the energy function corresponding to a cost function of an optimization problem, obtaining the equilibrium state for low enough temperature is already computationally hard.

Indeed, the Monte Carlo method mentioned above, for systems under low temperature, tends to have exponentially long relaxation times. In a way, this could be seen as a computational analogue of the Arrhenius law, which states that the relaxation time τ_{rel} in the low-temperature limit scales as

$$\tau_{\mathrm{rel}} \sim e^{\beta \Delta E^\star}. \tag{3.4}$$

Here, $\Delta E^\star$ is the characteristic energy scale of the energy barrier among the ground state configurations, which does not depend on the system size or temperature. This scaling does not hold when there is no characteristic energy scale of the energy barrier independent of system size and is a heuristic law.

One idea to overcome the slow relaxation in the low-temperature region is *annealing*. Since sampling the Boltzmann distribution for high temperature is easy (when $\beta = 0$, it is simply the uniform distribution), gradually decreasing the temperature starting from Monte Carlo sampling with high temperature may help. Intuitively, if the Boltzmann distribution does not change too drastically when the temperature is lowered, annealing should enable us to reach low temperature (therefore, phase transitions are common obstacles for SA). This intuition is backed up by the convergence result proven in [66, 67] which states that if the temperature T is decreased slower than

$$T(t) \propto \frac{N}{\log(1+t)}. \tag{3.5}$$

Since this bound is general, and is applicable to *all* systems with N Ising spins, it naturally yields an exponential upper bound on the computation time, consistent with the NP hardness assumption. If we want the temperature to be lower than a certain value ϵ, the necessary total time is upper bounded by $t_{\mathrm{fin}} \leq O(e^{N/\epsilon})$, which is exponential in the system size N.

As we written in section 2.2.2, dist-NP problems with natural distribution usually have a spin-glass phase transition at a finite temperature, and it is this phase transition which slows down SA.

Quantum Annealing

Quantum Annealing was first introduced by T. Kadowaki and H. Nishimori [51] for a quantum extension of SA.

Let us consider the situation where we want to compute the IsingSpinGlass problem defined in section 2.2.2 as an optimization problem. The original problem was to find a classical spin configuration for all $\sigma_i \in \{\pm 1\}$, which does not change its nature when rewritten in the quantum form

$$\hat{H}_{\mathrm{P}} = -\sum_{i<j} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z - \sum_i h_i \hat{\sigma}_i^z, \tag{3.6}$$

where $\hat{\sigma}_i^z$ denotes the Pauli-z matrix for the ith spin now. We then add a transverse field $\hat{H}_{\mathrm{T}} = -\sum_i \hat{\sigma}_i^x$ where $\hat{\sigma}_i^x$ is the Puli-x matrix for the ith spin, i.e.,

$$\hat{H}(\lambda) := (1-\lambda)\hat{H}_{\mathrm{P}} + \lambda \hat{H}_{\mathrm{T}}$$

$$= -(1-\lambda)\left(\sum_{i<j} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z - \sum_i h_i\hat{\sigma}_i^z\right) - \lambda\sum_i \hat{\sigma}_i^x. \tag{3.7}$$

The resulting Hamiltonian will have the same ground state as the original one at $\lambda = 0$ and becomes a natural quantum extension of SA. In the limit of strong transverse field ($\lambda = 1$), the ground state is the product state of all spins pointing in the x-direction, which could be written as $|+++\ldots+\rangle$. This is also the uniform superposition of *all* of the z-basis states,

$$|++\ldots+\rangle = \frac{1}{\sqrt{2^N}}\sum_{\boldsymbol{\sigma}\in\{|0\rangle,|1\rangle\}^{\otimes N}} |\boldsymbol{\sigma}\rangle, \tag{3.8}$$

naturally corresponding to the uniform distribution for $T \to \infty$ for SA. Since the ground states of both $\lambda = 0$ and $\lambda = 1$ are the same as the equilibrium states of SA at $T = 0$ and $T = \infty$, QA focuses on the $T = 0$ limit with varying λ. Starting from $\lambda = 1$ and gradually changing the annealing parameter λ is expected to have an analogous dynamics to that of SA. This is the basic idea of QA, and as we saw in the previous section, the quantum adiabatic theorem guarantees that the above expectation does hold if the annealing speed is slow enough i.e. $t_{\mathrm{fin}} \propto \Delta E_{\mathrm{min}}^{-2}$.

Thus, QA could be understood as a natural restriction of AQC to the form of Eq.(3.7). Another interpretation of QA is that it is the quantum extension of SA, utilizing quantum fluctuations instead of thermal fluctuations. In SA, the high-temperature limit could be seen as a classical probabilistic superposition of all configurations. As we decrease the temperature, the probabilistic superposition is suppressed, with less thermal fluctuation. Now in QA, since the temperature is fixed to $T = 0$, there is no thermal fluctuation, and quantum fluctuation induced by the transverse field plays the same role.

Summing up the above explanation, we can formalize QA as follows.

Definition 16. (Stoquastic)[*4] Quantum Annealing

1. Construct a spin Hamiltonian $\hat{H}_{\mathrm{P}}$ only using $\hat{\sigma}^z$ operators. $\hat{H}_{\mathrm{P}}$ should be the cost function of a combinatorial optimization problem.
2. Lower the temperature, and apply strong transverse field $\sum_i \hat{\sigma}_i^x$.
3. Gradually reduce the strength of the transverse field.
4. When the transverse field is turned off, measure the spins in the z-direction.

Although general results on the minimum energy gap ΔE_{min} are beyond our reach of rigorous proof, we have a general result on how unlikely it is to have a vanishing energy gap $\Delta E_{\mathrm{min}} = 0$.

Theorem 4. (No Crossing Theorem). A $d(<\infty)$ dimensional Hamiltonian with l different eigenvalues with $g_i (i = 1, 2, \ldots, l)$ degeneracies, has $d^2 + l - \sum_i g_i^2$ degrees of freedom.

If we think of a d dimensional Hermitian matrix in general, there are $d(d-1)/2$ off-diagonal elements each having two degrees of freedom (since they are complex), and the d diagonal elements have 1 degree of freedom (since they are real), adding up to d^2. This is the case where we set $l = d$ and $g_i \equiv 1$ for the above theorem. The theorem tells us that if there is degeneracy, the degree of freedom decreases. For instance, if there is just one degeneracy, we will have $l = d - 1$ and $g_i = 1$ except for one $g_0 = 2$, the resulting degrees of freedom will be $d^2 - 3$. This means that within the space of *all* possible Hamiltonians acting on a d dimensional Hilbert space, which turns out to be d^2

[*4] The term stoquastic will be defined in the following section.

dimensional, the subspace of the Hamiltonians with degeneracy is considerably smaller, with only dimension $d^2 - 3$. This suggests that the QA Hamiltonian Eq.(3.7) is *unlikely* to have degeneracies, except for points $\lambda = 0$ and $\lambda = 1$ where all the terms become commutative and the Hamiltonian can no longer be seen as typical.

The above argument for the QA Hamiltonian's non-degeneracy incorporating the No-crossing theorem is heuristic. There are known adiabatic cycles with crossings [68], but as far as for QA Hamiltonians, no natural-arising crossings are known for finite systems. Also, the theorem only applies to finite dimensional systems. This means in the thermodynamic limit $N \to \infty$, the energy gap may vanish at a quantum phase transition point.

Quantum phase transitions are sudden changes in the ground state of quantum systems. Similarly to the classical phase transitions at finite temperatures, we can classify quantum transitions either being *first order* or *second order* (continuous), by seeing discontinuity of physical observables. It is *generically* known that first order and second order quantum phase transitions have vanishing energy gaps at the transition point, which decrease exponentially and polynomially (power-law) to the system size respectively [69]. However, this correspondence is by no means a rigorous one, and counterexamples are also known for either side [70]. That being said, the existence of a first-order phase transition is still a good reference of knowing that it is likely to have an exponentially long annealing time, and is considered in many studies [71, 72].

It is also known that the first order transitions are hard for classical systems as well since local update search takes exponential time to find the other phase in general. However, the Wang-Landau algorithm is known to be very efficient for sampling classical systems which undergo first order phase transitions which could be regarded *easy* in some sense[*10] [73]. We will also see that a second order RSB will slow down the best algorithm in a certain NP complete problem in chapter 5, in the classical setting. Thus it is hard to have general arguments about first/second order transitions and their algorithmic barrier in the classical case.

3.2.2 Real-world Implementation and Stoquasticity

The D-wave Machine

The Canadian company D-wave systems Inc. announced they successfully build a QA devise in 2011 [74]. Since then, there were numerous controversies on if the announcement is true. It was shown in 2015 [75] that the 2000 qubit quantum annealer indeed displayed statistically similar results to the quantum Monte Carlo (QMC) simulations (explained in chapter 4). Furthermore, simulations assuming the qubits to be classical spins had statistically different results with the experimental demonstration. This result is thought to be evidence that the announced quantum annealer is indeed a physical implementation of QA. Furthermore, the paper showed that the physical quantum annealer was about 10^8 times faster compared to the QMC simulations. However, there are caveats to this result.

First of all, the problems that the quantum annealer tackled for the demonstration was designed in favour of the annealing device. While the most general form of QA is written in the form of Eq.(3.7), with arbitrary interaction for arbitrary pair of qubits, the quantum annealer device designed by D-wave has a particular structure called *chimera graph*, where only interactions on this graph are allowed. The test problems were designed to finely fit

[*10] The Wang-Landau algorithm tries to sample configurations to have uniform distribution over energy. This enables the algorithm to go "deeper into the phase" once it finds a low energy state, e.g., the all up state in the (fully-connected) three-body Ising model among few spins will "percolate" in order to search lower energy states. On the other hand, if the low energy state requires some *global* structure and is hard to find, the Wang-Landau algorithm seems to be powerless.

into the chimera graph structure, and whether problems in other forms, which embedding to the chimera graph is necessary, can benefit from the annealer remains unsettled.

Secondly, while a constant factor of 10^8 can be of enormous benefit for practical purposes, it does not imply quantum advantage we explained in section 2.3.2. Rather, the very fact of classical computers being able to simulate the quantum annealers hints that the device does not have quantum advantage at all in the sense it does not have better asymptotic scaling. This gave rise to a further controversy on what should the term *quantum computer* should address, since the company called the quantum annealer as the "world's first quantum computer"[76]. We will discuss this point in the following section.

Stoquasticity and Non-stoquastic Terms

We show again the general form of QA Eq.(3.7) here, as

$$\begin{aligned}\hat{H}(\lambda) &:= (1-\lambda)\hat{H}_{\mathrm{P}} + \lambda\hat{H}_{\mathrm{T}} \\ &= -(1-\lambda)\left(\sum_{i<j} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z - \sum_i h_i\hat{\sigma}_i^z\right) - \lambda\sum_i \hat{\sigma}_i^x.\end{aligned}$$

A very important feature of the above Hamiltonian is that it is simulable by the QMC method. While we leave the detailed explanation of QMC for the next chapter, we should emphasize that not *all* quantum Hamiltonians are simulable by QMC.

Although the class of *all* quantum systems which are simulable by some QMC is out of our current reach, we have a reasonable subclass which always allows QMC methods, called *stoquastic* Hamiltonians. The term "stoquastic" is (obviously) a coined word combining "stochastic" and "quantum", implying that the Hamiltonian has both quantum and classical probabilistic aspects. Let us first define the concept accordingly to [77].

Definition 17. Stoquastic Hamiltonian
A Hamiltonian $\hat{H}$ of an N qubit system is said to be *stoquastic* when the Hamiltonian could be decomposed into terms $\hat{H} = \sum_i \hat{h}_i$ with all of the matrix terms in the z-basis[*11] are real and non-posisitve, i.e.,

$$\langle\boldsymbol{\sigma}|\hat{h}_i|\boldsymbol{\sigma}'\rangle \leq 0 \quad \forall i \;\; \forall\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \{0,1\}^{\otimes N}. \tag{3.9}$$

All stoquastic Hamiltonians are simulable using QMC[*12] without facing the "negative sign" problem, which we will see shortly in the next chapter. An important observation is that the QA Hamiltonian Eq.(3.7) fits into this definition of stoquastic Hamiltonians. As we have seen in section 3.1, if we consider a more general Hamiltonian than QA, the framework becomes capable of computing BQP-complete problems. Notice that proving equivalence mentioned in section 3.1 relied on terms as $\hat{U}_l \otimes |011\rangle\langle 001|_{\mathrm{clock}}$ which is *non*stoquastic in general. Thus, a polynomial time AQC with *stoquastic* Hamiltonians becomes a natural restriction of AQC and is currently being studied both in the computational complexity field and physics. Let us call the class of problems solvable by the stoquastically restricted AQC in polynomial time[*13] as BstoqP for **"Bounded-error Stoquastic Polynomial"**[*14].

[*11] Obviously, equivalent classes of Hamiltonians could be defined using other local bases. Defining stoquasticity through *any* possible basis becomes problematic, since the energy eigenstate basis will trivially diagonalize the Hamiltonian with all off-diagonal elements 0.
[*12] This should not be confused with if it is polynomial time simulable.
[*13] Of course, allowing bounded error as in classes like BPP or BQP.
[*14] An equivalent complexity class is defined in [78], as QADI-SG, but we will not use this terminology for consistency with other classes in this thesis.

From the computational complexity viewpoint, a *dichotomy result* is known for quantum Hamiltonians [79]. It is proved that all *general* quantum Hamiltonian for qubits could be classified into groups of the either four.

- The most "trivial" Hamiltonians where there are no interactions but only local fields. For this class, the problem of determining whether the ground state energy is below some value, which we call LOCAL HAMILTONIAN[*15], is in P.
- The "classical" Hamiltonians where all terms are simultaneously diagonalizable. For this class, the problem LOCAL HAMILTONIAN is NP-complete.
- The "stoquastic" Hamiltonians where all the terms could be written in the form of Eq.(3.7) simultaneously by a rotation. For this class, LOCAL HAMILTONIAN becomes StoqMA-complete.
- The "truly quantum" Hamiltonians where none of the above applies. In this case, LOCAL HAMILTONIAN is QMA-complete.

We have not defined in this thesis the complexity classes stoqMA and QMA. Here, it will be good enough to know that they are the stoquastic and quantum analogue of NP respectively. The important point is that a complexity class could be naturally defined, and a problem regarding the ground state of the Hamiltonian of QA (Eq.3.7) becomes complete for that class. Thus if we believe that NP≠stoqMA, just like we believe P≠NP, then QA should have quantum advantage for *some problems*. On the other hand, while P and NP were believed to be different for many reasons as we saw in the previous chapter, we have only little arguments for separating NP and stoqMA. Whether if QA really has a quantum advantage (i.e. whether if it is polynomially simulable using classical computers, or equivalently, if BstoqP=BPP) will be discussed in the next chapter.

So far, we have only considered QA as a stoquastically restricted AQC protocol. This is not necessarily the only interesting case, since the original idea of QA was just to have "the quantum analogue of SA". Changing the Hamiltonian fast ignoring the adiabatic condition is one direction. In this case, the state does not remain in the instantaneous ground state, and a Landau-Zener type transition occurs[80, 81]. This is called non-adiabatic transitions to excited states. For instance, there have been studies on QA with multiple non-adiabatic transitions, which the state ends up in the ground state [82]. There is also research direction ignoring the adiabatic condition and using QA as a way of sampling approximate solutions [83].

Another way of extending QA is to use *non-stoquastic* terms. A pioneering work in this direction was done in [72], where an antiferromagnetic term in the x-direction $\sum_{i<j} \hat{\sigma}_i^x \hat{\sigma}_j^x$ was introduced to reduce the intensity of a quantum phase transition from first order to second order. While the computational value of this non-stoquastic term is currently unknown from the complexity point of view[*16], it is numerically observed that the antiferromagnetic term tends to widen the energy gap [84].

[*15] To be precise, it needs to be a *promise* problem where it is guaranteed that the ground state energy is either above a or below b with a polynomially bounded gap $b - a \geq 1/\mathrm{poly}(N)$. This applies to all of the problems listed here. For details, see [79].

[*16] It is only proved that the LOCALHAMILTONIAN problem becomes QMA-complete when we allow terms such as $\sum_{i<j} \alpha_{ij} \hat{\sigma}_i^x \hat{\sigma}_j^x$, with tunable interactions. Thus, the uniform antiferromagnetic term only by itself does not seem sufficient for completeness (though not proved to be insufficient either).

Chapter 4

Stoquastic Quantum Annealing and its Simulability

In the previous chapter, we have seen that the most studied (and also the only physically implemented) form of QA is stoquastic. The current understanding of stoquastic Hamiltonians is that they form a natural complexity class, which lies somewhere between classical computation and universal quantum computation. While the possibility of where the complexity class BstoqP actually lies is totally open, it is true that the stoquastic property of the QA Hamiltonian lets us simulate QA to some extent.

In this chapter, we will discuss how the simulation is done, and see possibilities of further determining the computational complexity position of stoquastic Hamiltonians. We especially focus on the possibility where QMC methods might allow us to show that stoquastic QA is always polynomially simulable in the last section.

4.1 The Quantum Monte Carlo Method

The Quantum Monte Carlo (QMC) method is a probabilistic algorithm for simulating quantum systems. There are numbers of different ways this could be done, but the basic idea of "transforming a quantum system into a classical one" is common among all of them. In this section, we will explain the path integral Monte Carlo which is the most basic type of QMC, and also the stochastic series expansion (SSE) which is a more sophisticated algorithm which we use in the following section.

4.1.1 Basics of Monte Carlo Methods

The term "Monte Carlo method" refers to the broad class of probabilistic algorithms where the accuracy of the algorithm increases in time (usually exponentially). While there are also Monte Carlo algorithms for algebraic problems like PRIMES, in this thesis we will only use the Monte Carlo method as a sampling method (applying it for simulations). In this section, we will see the basic idea of Markov chain Monte Carlo (MCMC) algorithms, which is a subclass of Monte Carlo algorithms.

It is the easiest to consider sampling states $\{\boldsymbol{\sigma}\}$ of a classical system with discrete degree of freedom, from the Boltzmann distribution as in Eq.(3.3). The difficulty comes from the fact that for a thermodynamically sound system, the number of all states Ω grows exponentially to the system size N. Thus, explicitly calculating the partition function Z (which is the normalization constant in this context) will require exponential time if we naively add up the sums.

We consider a Markov process among the Ω states $\{\boldsymbol{\sigma}_a\}_{a=1}^{\Omega}$. Starting from any state, the next state is determined probabilistically, only depending on the current state. For example, if the state at time t was $\boldsymbol{\sigma}_b$, the probability of moving to $\boldsymbol{\sigma}_a$ is a constant which we denote P_{ab}. If we express the probability of being in state $\boldsymbol{\sigma}_a$ at time t as $\pi_a(t)$, we can write

$$\begin{pmatrix} \pi_1(t+1) \\ \pi_2(t+1) \\ \vdots \\ \pi_\Omega(t+1) \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1\Omega} \\ P_{21} & P_{22} & \cdots & P_{2\Omega} \\ \vdots & \vdots & \ddots & \vdots \\ P_{\Omega 1} & P_{\Omega 2} & \cdots & P_{\Omega\Omega} \end{pmatrix} \begin{pmatrix} \pi_1(t) \\ \pi_2(t) \\ \vdots \\ \pi_\Omega(t) \end{pmatrix}. \tag{4.1}$$

This is called the master equation of a Markov chain, and the matrix is called the transition matrix. If we simply denote the matrix as P and the vector as $\boldsymbol{\pi}(t)$, we get the form

$$\boldsymbol{\pi}(t) = P^t \boldsymbol{\pi}(0). \tag{4.2}$$

It is proved that the probability distribution $\boldsymbol{\pi}(t)$ will converge to the equilibrium distribution $\boldsymbol{\pi}^{(\mathrm{eq})}$ from any initial distribution $\boldsymbol{\pi}(0)$ if the transition matrix P satisfies the following conditions:

1. $\boldsymbol{\pi}^{(\mathrm{eq})}$ is a right-eigen vector of P with eigen value 1.
2. $\exists M \in \mathbb{N},\ \ 1 \le \forall i, j \le \Omega,\ m > M \rightarrow (P^m)_{ij} > 0$.

The existence of a distribution $\boldsymbol{\pi}^{\mathrm{eq}}$ in Condition 1 is guaranteed from the Perron-Frobenius theorem [85], as long as if P is a valid transition matrix (i.e. $\forall b\ \sum_a P_{ab} = 1,\ P_{ab} \ge 0$). Condition 2 represents what is called *irreducibility* and *aperiodicity* of a Markov chain[*1], guaranteeing that the Markov chain indeed mixes the states. Without this condition, the uniqueness of $\boldsymbol{\pi}^{\mathrm{eq}}$ does not hold. We will not go further on this condition since it will be always satisfied in our following studies. The first condition is also known as the *balance condition* stating that $\sum_b \pi_b^{(\mathrm{eq})} P_{ab} = \pi_a^{(\mathrm{eq})}$. For simplicity, the tighter condition of *detailed balance*

$$\pi_a^{(\mathrm{eq})} P_{ba} = \pi_b^{(\mathrm{eq})} P_{ab}, \tag{4.3}$$

is often considered[*2], meaning that the probability flow from state a to b is equal to that from b to a in equilibrium. The heat bath method

$$P_{ab} = \frac{\pi_a^{(\mathrm{eq})}}{\pi_a^{(\mathrm{eq})} + \pi_b^{(\mathrm{eq})}}, \tag{4.4}$$

and the Metropolis method

$$P_{ab} = \min[\ 1\ ,\ \ \frac{\pi_a^{(\mathrm{eq})}}{\pi_b^{(\mathrm{eq})}}\ ], \tag{4.5}$$

are two famous frameworks of constructing a transition probability satisfying detailed balance, which only requires the *ratio* of the equilibrium distribution (and thus without knowing the partition function Z for the statistical mechanics case) for calculation.

4.1.2 Path Integral Monte Carlo

The idea of path integral Monte Carlo (PIMC) is to transform a quantum (stoquastic) Hamiltonian into a classical one. Let's consider a quantum spin 1/2 system with the

[*1] In the computational physics literature, these conditions may also be refferred as "ergodicity".
[*2] There has been recent progress on how breaking the detailed balance condition may enhance equilibration of the Markov Chain [87, 88].

Hamiltonian in the form

$$\hat{H} = \hat{\mathcal{F}}(\{\hat{\sigma}_i^z\}) - \Gamma \sum_{i=1}^{N} \hat{\sigma}_i^x \tag{4.6}$$

where $\hat{\sigma}_i^\alpha$ denotes the Pauli matrix $\hat{\sigma}_i^\alpha$ acting on the ith spin, i.e.,

$$\hat{\sigma}^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \hat{\sigma}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{4.7}$$

and $\hat{\mathcal{F}}$ is an arbitrary function of $\{\hat{\sigma}_i^z\}$, e.g. $\sum_{i,j} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z$ or $\frac{1}{N^{p-1}}(\sum_i \hat{\sigma}_i^z)^p$ etc.

According to quantum statistical mechanics, the matrix of a quantum system with Hamiltonian $\hat{H}$ at inverse temperature β could be expressed as

$$\hat{\rho} := \frac{1}{Z} e^{-\beta \hat{H}}, \quad Z := \mathrm{Tr}[e^{-\beta \hat{H}}], \tag{4.8}$$

and the expectation value of an observable $\hat{A}$ is calculated by

$$\langle \hat{A} \rangle = \mathrm{Tr}[\hat{\rho}\hat{A}]. \tag{4.9}$$

The central difficulty of calculating physical quantities from the above framework arises from calculating Z. Especially for quantum systems, the Hamiltonian generally has non-commuting terms (for the current example, $[\hat{\sigma}_i^z, \hat{\sigma}_i^x] = 2i\hat{\sigma}_i^y \neq 0$), which complicates the situation, since for non-commuting matrices $\hat{A}$ and $\hat{B}$, $e^{\hat{A}+\hat{B}} \neq e^{\hat{A}}e^{\hat{B}}$ in general.

The strategy of PIMC is to use the Lie product formula

$$e^{\hat{A}+\hat{B}} = \lim_{L\to\infty} \left(e^{\frac{\hat{A}}{L}} e^{\frac{\hat{B}}{L}} \right)^L, \tag{4.10}$$

to calculate the density matrix. By expanding the trace as sums over the z-basis states $\{|\boldsymbol{\sigma}\rangle\}$*3, we obtain

$$\begin{aligned}
Z &= \mathrm{Tr}[e^{-\beta\hat{H}}] \\
&= \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | e^{-\beta\hat{\mathcal{F}}(\{\hat{\sigma}_i^z\}) + \beta\Gamma \sum \hat{\sigma}_i^x} | \boldsymbol{\sigma} \rangle \\
&= \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | \lim_{L\to\infty} (e^{-\frac{\beta}{L}\hat{\mathcal{F}}(\{\hat{\sigma}_i^z\})} e^{\frac{\beta\Gamma}{L}\sum \hat{\sigma}_i^x})^L | \boldsymbol{\sigma} \rangle \\
&= \lim_{L\to\infty} \sum_{\boldsymbol{\sigma}(0)} \langle \boldsymbol{\sigma}(0) | \prod_{t=1}^{L} \left\{ e^{-\frac{\beta}{L}\hat{\mathcal{F}}(\{\hat{\sigma}_i^z\})} \ \Big(\sum_{\boldsymbol{\sigma}(t)} |\boldsymbol{\sigma}(t)\rangle\langle\boldsymbol{\sigma}(t)|\Big) \ e^{\frac{\beta\Gamma}{L}\sum\hat{\sigma}_i^x} \right\} |\boldsymbol{\sigma}(0)\rangle \\
&= \lim_{L\to\infty} \sum_{\boldsymbol{\sigma}(1),\boldsymbol{\sigma}(2),\ldots,\boldsymbol{\sigma}(L)} e^{-\frac{\beta}{L}\sum_t \mathcal{F}(\boldsymbol{\sigma}(t))} \Big(\prod_{t=1}^{L} \langle \boldsymbol{\sigma}(t) | e^{\frac{\beta\Gamma}{L}\sum\hat{\sigma}_i^x} | \boldsymbol{\sigma}(t+1) \rangle \Big).
\end{aligned}$$

Here $L+1 \equiv 1$ as in periodic boundary conditions. $\mathcal{F}$ without the hat at the last line means it is just a classical function now. Notice that the z basis $|\boldsymbol{\sigma}\rangle$ passes through $\mathcal{F}$. We can use a simple relation

$$e^{\alpha\hat{\sigma}^x} = (\cosh\alpha)\hat{1} + (\sinh\alpha)\hat{\sigma}^x \tag{4.11}$$

*3 Although we have introduced the z-basis as $|0\rangle$ and $|1\rangle$ in chapter 2, here we will use $|\sigma\rangle$, $\sigma \in \{\pm 1\}$ for convenience. This representation is more commonly used in physics.

for evaluating the term $\langle \boldsymbol{\sigma}(t)|\exp(\frac{\beta\Gamma}{L}\sum\hat{\sigma}_i^x)|\boldsymbol{\sigma}(t+1)\rangle$, which becomes

$$\begin{aligned}
\langle \boldsymbol{\sigma}(t)|e^{\frac{\beta\Gamma}{L}\sum_{i=1}^N \hat{\sigma}_i^x}|\boldsymbol{\sigma}(t+1)\rangle &= \prod_{i=1}^N \langle \sigma_i(t)|\exp\left(\frac{\beta\Gamma}{L}\hat{\sigma}_i^x\right)|\sigma_i(t+1)\rangle \\
&= \prod_{i=1}^N \left(\delta_{\sigma_i(t),\sigma_i(t+1)}\cosh\frac{\beta\Gamma}{L} + \delta_{\sigma_i(t),-\sigma_i(t+1)}\sinh\frac{\beta\Gamma}{L}\right) \\
&= \sqrt{\frac{1}{2}\sinh\frac{2\beta\Gamma}{L}}^N \prod_{i=1}^N \left(\tanh\frac{\beta\Gamma}{L}\right)^{-\frac{1}{2}\sigma_i(t)\sigma_i(t+1)} \\
&= \exp\left[\frac{N}{2}\log(\frac{1}{2}\sinh\frac{2\beta\Gamma}{L}) - \frac{1}{2}\log(\tanh\frac{\beta\Gamma}{L})\sum_{i=1}^N \sigma_i(t)\sigma_i(t+1)\right].
\end{aligned}$$

If we denote the $\boldsymbol{\sigma}$ independent term $\frac{1}{2}\log\frac{1}{2}\sinh\frac{2\Gamma}{L}$ as C, the partition function could be written as

$$Z = \lim_{L\to\infty}\sum_{\{\boldsymbol{\sigma}(t)\}}\exp\left(-\frac{\beta}{L}\sum_{t=1}^L \mathcal{F}(\boldsymbol{\sigma}(t)) + NLC - \frac{1}{2}\log\left(\tanh\frac{\beta\Gamma}{L}\right)\sum_{t=1}^L\sum_{i=1}^N \sigma_i(t)\sigma_i(t+1)\right), \tag{4.12}$$

which could be seen as a partition function of a *classical* Hamiltonian

$$H_{\mathrm{ST}} = \frac{1}{L}\sum_t \mathcal{F}(\boldsymbol{\sigma}(t)) + \frac{1}{2\beta}\log\left(\tanh\frac{\beta\Gamma}{L}\right)\sum_{t=1}^L\sum_{i=1}^N \sigma_i(t)\sigma_i(t+1) + NLC, \tag{4.13}$$

with classical Ising spins $\{\sigma_i(t)\}_{i=1,2,\ldots,N}^{t=1,2,\ldots,L}$. The above calculation based on the Lie formula is called *Suzuki-Trotter expansion*[*4].

Now that an equivalent classical Hamiltonian for the original stoquastic Hamiltonian is found, any classical Monte Carlo algorithm (e.g. the Metropolis algorithm) can be used. The constant term NLC could be completely ignored in this case. Note that the resulting classical Hamiltonian has a temperature dependent term which is rather pathological.

A caveat to this approach is that the equivalence is only accurate in the $L\to\infty$ limit. In general, the Monte Carlo simulation is possible only for finite systems. Thus an extrapolation after setting L to a finite-value is necessary, where the finite valued L is called the *Trotter number*. The system represented by the classical Hamiltonian H_{ST} could be seen as the original system (expressed by $\hat{H}$) extending in an additional dimension with the periodic boundary condition, labelled by $t = 1, 2, \ldots, L$. This direction is called the *Trotter direction*, and we can see that the classical spins have ferromagnetic interactions along this direction.

An important point of the Suzuki-Trotter expansion is that stoquastic Hamiltonians are always reducible to classical Hamiltonians in the same way[*5]. The arbitrariness of $\mathcal{F}$ corresponds to the arbitrariness of diagonal elements for the definition of stoquastic Hamiltonians[*6]. If there are non-stoquastic terms in the Hamiltonian, e.g. $\alpha\hat{\sigma}_i^x\hat{\sigma}_j^x$, the rewriting corresponding to Eq.(4.11) may yield results with *negative* signs. The reason

---

[*4] M. Suzuki [86] was the first to point out that the Lie formula could be used for numerical simulation by cutting off the supposedly infinite L.

[*5] The transverse field Ising model in the form of Eq.(3.7) is known to be "stoquastic-complete" [89]. This means that any stoquastic Hamiltonian can be efficiently simulated by a Hamiltonian expressed by Eq.(3.7) with polynomially many terms, in terms of the energy spectrum.

[*6] It should be noted that the function of $\mathcal{F}$ must be *efficiently computable* for PIMC to work. More

why the PIMC worked for the stoquastic Hamiltonian Eq.(4.6) is because the uniform transverse field $\sum_i \hat{\sigma}_i^x$ itself will never give such negative signs.

4.1.3 Stochastic Series Expansion

The PIMC method required extrapolation to the $L \to \infty$ limit. There are a number of ways which directly samples the system corresponding to the $L \to \infty$ limit [90, 91]. Stochastic series expansion (SSE) algorithm is also a Quantum Monte Carlo method which effectively corresponds to the $L \to \infty$ limit. The advantage of using SSE over other $L \to \infty$ methods is that SSE only samples among discrete variables, which is practically faster than computing continuous variables*7. Furthermore, the Swendsen-Wang type update adopted in SSE avoids slowdowns which loop algorithms face when dealing with local magnetic fields [92]. We will explain the SSE in detail, since we use this method later in this chapter to see the robustness of "stoquastic advantage", and also to see how stoquasticity is dealt in this method.

Basics

Similarly to the introduction of PIMC, considering how to express the density matrix $Z = \mathrm{Tr}[e^{-\beta \hat{H}}]$ in terms of classical probability serves as a good starting point. The essential difference with PIMC is that instead of using the Suzuki-Trotter expansion, we rewrite the density matrix using the Taylor expansion as

$$Z = \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | \sum_{n=0}^{\infty} \frac{\beta^n}{n!} (-\hat{H})^n | \boldsymbol{\sigma} \rangle. \tag{4.14}$$

Rewriting the trace as a simple sum over basis states is the same as PIMC. The basis should be chosen carefully, due to reasons which will be apparent soon. The next rewriting is crucial for executing SSE, and that is writing $\hat{H}$ as $\hat{H} = -\sum_k \hat{W}_k$, where we assume that all of the eigenvalues of $\hat{W}_k$ are non-negative. This enables us to get rid of the negative sign in the equation as

$$Z = \sum_{n=0}^{\infty} \sum_{\boldsymbol{\sigma}} \frac{\beta^n}{n!} \langle \boldsymbol{\sigma} | \big(\sum_k \hat{W}_k \big)^n | \boldsymbol{\sigma} \rangle = \sum_{n=0}^{\infty} \sum_{\boldsymbol{\sigma}} \sum_{k(l)} \frac{\beta^n}{n!} \langle \boldsymbol{\sigma} | \prod_{l=1}^{n} \hat{W}_{k(l)} | \boldsymbol{\sigma} \rangle. \tag{4.15}$$

Now, estimating the partition function Z is reduced to a sampling problem over the basis state $|\boldsymbol{\sigma}\rangle$ and the operators $\hat{W}_k$ (in this section, we will call them as "configurations"). Since we required that all the eigenvalues are non-negative, all configurations have non-negative weight accordingly to the operators. The basis should be chosen so that the operators working on them wouldn't mess up the states.

The weight of every operator alignments cannot grow infinitely since there is a $1/n!$ factor, and we can assume that the typical value of n from a sampled configuration is concentrated around its average value *8. If we go well above the average value of n, the

precisely, it can only contain at most polynomial numbers of terms as a polynomial. This suggests a further restriction within stoquastic Hamiltonians (namely, locality) for QMC approaches to be valid. On the other hand, when the function $\mathcal{F}$ requires exponential time for computation, there seems to be no way of constructing such a Hamiltonian for the quantum annealer, in polynomial time.

*7 In practice, we cannot handle *true* real numbers, since that will require infinite memory. Here, we are referring to the fact that usual double-precision floating-point type of "real numbers" takes longer time for an operation compared to integer type.

*8 We will soon see that the value of n represents the energy of the system. This means that if we are dealing with a regular system with bounded energy, the argument of n being concentrated is valid.

weight of that configuration would be exponentially small. Thus we can have a cut-off value L for n. Then, we can rewrite the equation so that there are always L operators multiplied. This is because we can think that there are $L-n$ "null-operators" $\hat{1}$ and have a special $k=\kappa$ assigned to it. We obtain

$$Z = \sum_{n=0}^{L} \sum_{\boldsymbol{\sigma}} \sum_{k(l)} \frac{\beta^n}{n!} \langle \boldsymbol{\sigma} | \prod_{l=1}^{n} \hat{W}_{k(l)} | \boldsymbol{\sigma} \rangle = \sum_{\boldsymbol{\sigma}} \sum_{k'(l)} \frac{\beta^n}{n!} \frac{1}{{}_L C_n} \langle \boldsymbol{\sigma} | \prod_{l=1}^{L} \hat{W}_{k'(l)} | \boldsymbol{\sigma} \rangle \tag{4.16}$$

$$= \sum_{\boldsymbol{\sigma}} \sum_{k'(l)} \frac{\beta^n (L-n)!}{L!} \langle \boldsymbol{\sigma} | \prod_{l=1}^{L} \hat{W}_{k'(l)} | \boldsymbol{\sigma} \rangle. \tag{4.17}$$

Note that there is a degree of freedom where we get to choose which $L-n$ possitions we put in the null-operators. This should result in a ${}_L C_n$ factor, and we get the final equation.

As a result, we have a finite L and at the same time could be sure that the sampling is indistinguishable from an exact one as long as the number of operators n does not reach L. The line of operators $\{\hat{W}_{k'(l)}\}_{l=1}^{L}$ plays a similar role to the Trotter direction in PIMC, with periodic boundary condition. In fact, these two could be seen equivalent under certain transformation [93], and we will refer to the line of operators also as the Trotter direction.

Measuring Observables

To apply SSE to measuring an observable $\hat{A}$, we start from the equation $\langle \hat{A} \rangle = \mathrm{Tr}[e^{-\beta \hat{H}} \hat{A}]/Z$ and then rewrite it analogously. We obtain

$$\langle \hat{A} \rangle = \frac{\sum_{\boldsymbol{\sigma}} \sum_{k'(l)} \frac{\beta^n (L-n)!}{L!} \langle \boldsymbol{\sigma} | \prod_{l=1}^{n} \hat{W}_{k'(l)} \hat{A} | \boldsymbol{\sigma} \rangle}{\sum_{\boldsymbol{\sigma}} \sum_{k'(l)} \frac{\beta^n (L-n)!}{L!} \langle \boldsymbol{\sigma} | \prod_{l=1}^{n} \hat{W}_{k'(l)} | \boldsymbol{\sigma} \rangle}, \tag{4.18}$$

meaning that if the basis state is an eigenstate of the observable $\hat{A}$, then simply taking the average of the eigenvalues for the sampled configurations would yeild the desired value of observable $\hat{A}$. We will write this as $\langle A(\boldsymbol{\sigma}) \rangle_{\mathrm{config}} = \langle \hat{A} \rangle$. These observables are called *diagonal* observables.

Measuring observables which are non-commutative with the basis is more complicated. We first take the example of measuring the energy. By a simple rearrangement, we get

$$E = \langle \hat{H} \rangle = \frac{1}{Z} \mathrm{Tr}[e^{-\beta \hat{H}} \hat{H}] = \frac{1}{Z} \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | \sum_{n=0}^{\infty} \frac{\beta^n}{n!} (-\hat{H})^{n+1} (-1) | \boldsymbol{\sigma} \rangle$$

$$= \frac{1}{Z} \sum_{\boldsymbol{\sigma}} \langle \boldsymbol{\sigma} | \sum_{n=0}^{\infty} \frac{\beta^n}{n!} (-\hat{H})^n (-\frac{n}{\beta}) | \boldsymbol{\sigma} \rangle. \tag{4.19}$$

By comparing the result with Eq.(4.18), we see that taking the expectation value of n, the number of operators that are not null-operators, gives us the energy. In other words,

$$E = \langle \hat{H} \rangle = -\frac{\langle n \rangle_{\mathrm{config}}}{\beta}. \tag{4.20}$$

We can also calculate the specific heat similarly, and we obtain

$$C = \langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2 = \langle n^2 \rangle_{\mathrm{config}} - \langle n \rangle_{\mathrm{config}}^2 - \langle n \rangle_{\mathrm{config}}. \tag{4.21}$$

Other observables which are contained in the Hamiltonian (eg. $\langle \hat{H}_A \rangle$ such that $\hat{H} = \hat{H}_A + \hat{H}_B$) could be measured analogously, namely, by counting the number of operators corresponding to the term. By similar calculation, we obtain

$$E_A = \langle \hat{H}_A \rangle = -\frac{\langle n_A \rangle_{\mathrm{config}}}{\beta}, \tag{4.22}$$

where n_A denotes the number of operators corresponding to $\hat{H}_A$ in the configuration.

Example for an Ising Spin Glass

In this section, an example of SSE is presented using the Ising spin-glass model. We will consider the following stoquastic Hamiltonian.

$$\hat{H} = -\sum_{\langle ij \rangle} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z - \sum_i h_i \hat{\sigma}_i^z - \sum_i \Gamma_i \hat{\sigma}_i^x \tag{4.23}$$

In this model, all operators in the Hamiltonian are either a product of $\hat{\sigma}_i^z$ operators or just a single $\hat{\sigma}_i^x$, so using the z basis is convenient. All the operators will either leave the basis unchanged or just flip one qubit.

Other than the basis to use, there are two points which need to be considered. One is that we will have to rewrite this in the form $\hat{H} = -\sum_k \hat{W}_k$ as mentioned in the previous section, making sure that all the eigenvalues of $\hat{W}_k$ are non-negative. This could be done by adding adequate constants to each terms, as exemplified in the following.

$$J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z \to J_{ij} (\hat{\sigma}_i^z \hat{\sigma}_j^z \pm \hat{1}) \tag{4.24}$$

$$h_i \hat{\sigma}_i^z \to h_i (\hat{\sigma}_i^z \pm \hat{1}) \tag{4.25}$$

Whether if we add or subtract 1 depends on the sign of the coefficients. These will result in an extra constant term in the Hamiltonian, which does not make any physical difference*[9].

Let us now consider the effect of having biased weights as above. For example, $J_{ij}(\hat{\sigma}_i^z \hat{\sigma}_j^z \pm \hat{1})$ has 2 eigenvalues, namely $2|J_{ij}|$ and 0. This means that the configuration can have non-zero weight only when the operators work on states *consistent* with the signs of their coefficients. More precisely, $J_{ij}(\hat{\sigma}_i^z \hat{\sigma}_j^z \pm \hat{1})$, depending on if J_{ij} is positive or negative, can only be inserted in places where $|\sigma_i\rangle$ and $|\sigma_j\rangle$ have same/different values respectively. It is the same for $h_i(\hat{\sigma}_i^z \pm \hat{1})$, and they could be inserted only in places where the value of $|\sigma_i\rangle$ matches h_i's sign.

The second point we need to consider is the periodic boundary of Trotter direction. Stated more precisely, $|\sigma\rangle$ needs to be exactly the same except for scalar factors after applying the operators $\prod_{l=1}^{n} \hat{W}_{k'(l)}$ if we use an orthonormal basis (and let's assume that is the case). The $\hat{\sigma}_i^z$ operators cast no problem in this sense, and in this model, $\hat{\sigma}_i^x$ is the problem. If there are odd numbers of $\hat{\sigma}_i^x$ for a particular site i, the resulting state would differ with $|\sigma\rangle$ at site i and thus the entire weight would become 0. There are several possible ways to avoid this problem. One is that we just ignore configurations with odd $\hat{\sigma}_i^x$ while we naively sample. However, to do this, it is hard to choose the basis $|\sigma\rangle$ according to the equilibrium distribution and we will take another approach.

*[9] Since we only require the non-negativity of the eigenvalues, the constant which is added can be any value in principle as long as if it's greater than 1. It can also differ from site to site. While changing the constants in these ways *may* result in faster convergence of the Markov chain *in principle*, it has no physical meaning, and no literature is found in this direction. It appears to be an unnecessary complication without any physical meaning, to consider this degree of freedom.

Swendsen-Wang type cluster updates[94] are the usual solution to this problem, and it also has the advantage of making the equilibration faster. When there are external fields h_i we need a slight modification to the basic cluster update explained in [95]. The central idea is that we introduce another constant $\Gamma_i\hat{1}$. These constants will have exactly the same weight as $\Gamma_i\hat{\sigma}_i^x$, and the only difference is whether if they flip the spins or not.

First, local updates are carried out by putting in or pulling out operators, without changing the $\Gamma_i\hat{\sigma}_i^x$ operators. Here, operators $\Gamma_i\hat{1}$ will be inserted with the according probabilities (which is the same as $\Gamma_i\hat{\sigma}_i^x$ since they have the same weight) without flipping the spins, thus not violating the periodic boundary condition. Then, cluster updates are executed, where *clusters* in the configuration are defined, and get flipped with probability 1/2. Each cluster flips will be conducted so that the cluster flip will *not* change the weight of the configuration. If there are no external fields h_i, this is easy and we should just flip an entire region which is connected by $J_{ij}(\hat{\sigma}_i^z\hat{\sigma}_j^z \pm \hat{1})$ operators and is cut by either $\Gamma_i\hat{\sigma}_i^x$ or $\Gamma_i\hat{1}$. When flipping this cluster, the edges will change from $\hat{\sigma}_i^x$ to $\hat{1}$ or vice versa. Due to the periodic boundary, a cluster may be connected via the endpoints ($l = 1$ and $l = L$). In this case, the basis state $|\sigma\rangle$ should also be flipped at the sites which are included in the flipped clusters.

When there are external fields, there are also $h_i(\hat{\sigma}_i^z \pm \hat{1})$ operators. They can be inserted or removed just like other operators in the local update phase, in cluster updates however, they play a different role. If a cluster includes an $h_i(\hat{\sigma}_i^z \pm \hat{1})$ operator inside, the configuration will have 0 weight after the cluster is flipped. Thus that particular cluster with an $h_i(\hat{\sigma}_i^z \pm \hat{1})$ operator in it *cannot* be flipped. The resulting operation would be to flip all the clusters without $h_i(\hat{\sigma}_i^z \pm \hat{1})$ operators with probability 1/2, and unchange the others. Getting back to the original problem, this would always maintain the periodic boundary condition and we would never get a configuration with odd numbers of $\hat{\sigma}_i^x$ operators.

Stoquasticity

The example we considered so far for SSE had at most two-body interactions. This is not always the case if we consider stoquastic Hamiltonians in general. However, it should be clear how to generalize the above procedure to many-body interactions.

In fact, later on in section 4.3, we will consider an N-body interaction which is in the form $-|\boldsymbol{\sigma}\rangle\langle\boldsymbol{\sigma}|$. The basics are the same, that we insert/remove an interaction operator (another example could be $J_{ijk}(\hat{\sigma}_i^z\hat{\sigma}_j^z\hat{\sigma}_k^z \pm 1)$) so that it is consistent with the spin configuration, and regard them as a connected cluster. Thus, all Hamiltonians in the form of Eq.(4.6) could be put into the SSE formulation in principle. However, there is the same caveat as for PIMC. The function $\mathcal{F}$ must be polynomially computable in order to efficiently do the Monte Carlo computation. In the case of SSE, if there are exponentially many terms, the necessary L becomes exponentially large.

The Actual Sampling Protocol

In this subsection, we will explain the actual protocol for a Ising spin-glass system with examples. As we explained in the previous section, SSE is a framework wich samples "configurations" accordingly to Eq.(4.17)

$$Z = \sum_{\boldsymbol{\sigma}} \sum_{k'(l)} \frac{\beta^n (L-n)!}{L!} \langle\boldsymbol{\sigma}| \prod_{l=1}^{L} \hat{W}_{k'(l)} |\boldsymbol{\sigma}\rangle.$$

The configuration here is the list of L operators $\{\hat{W}_{k'(l)}\}_{l=1,2,\dots,L}$ together with the basis $|\boldsymbol{\sigma}\rangle$ which could be expressed as an N digit binary number. Let us continue to use the

Hamiltonian

$$\hat{H} = -\sum_{\langle ij \rangle} J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z - \sum_i h_i\hat{\sigma}_i^z - \sum_i \Gamma_i\hat{\sigma}_i^x,$$

for explaining the two phases of the actual sampling procedure, namely the local and global updates.

In the local update phase, we focus on one label of $l \in \{1, 2, \ldots, L\}$ at a time, and replace the operator $\hat{W}_{k'(l)}$ with a different one with appropriate prbability. For the current example, the operators we consider are

1. the null operator $\hat{1}$,
2. interaction operators $J_{ij}(\hat{\sigma}_i^z\hat{\sigma}_j^z \pm \hat{1})$,
3. transverse field operators $\Gamma_i\hat{\sigma}_i$,
4. transverse supporting operators $\Gamma_i\hat{1}_i$, and
5. local field operators $h_i(\hat{\sigma}_i^z \pm \hat{1})$.

The operator string $\{\hat{W}_{k'(l)}\}_{l=1}^L$ always needs to be consistent with the basis $|\boldsymbol{\sigma}\rangle$. It is convenient to imagine a diagram of $N \times L$ classical spins corresponding to the configuration (Fig.4.1). Here, the N spins at the edges are the same as $|\boldsymbol{\sigma}\rangle$, and the classical spins get flipped along the Trotter direction whenever encountering a transverse field operator. The interaction operators and local fields can only appear in places consistent with these classical spins, as depicted in the following diagram.

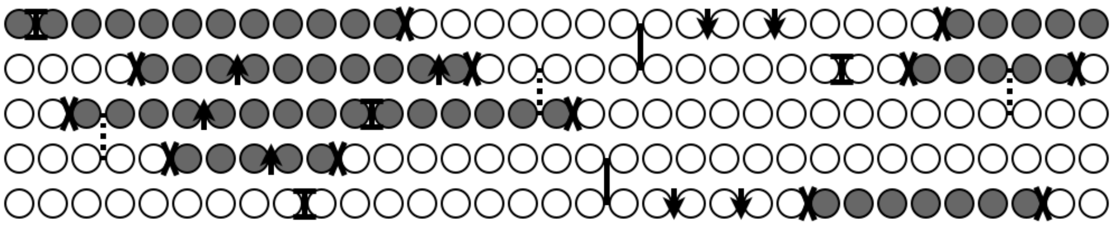

Fig. 4.1: An example of a configuration for a 1-dimentional Ising spin glass with 5 sites and L=31. The actual configuration only consists of the spin configuration at the edge and L operators in order. The white(down spins)/grey(up spins) circles are there only to make the picture clearer. I represents the transverse supporting operator $\Gamma_i\hat{1}_i$, X represents the transverse field oeprator $\Gamma_i\hat{\sigma}_i^x$, solid and dotted lines connecting the rows represnt interaction operators $J_{ij}(\hat{\sigma}_i^z\hat{\sigma}_j^z \pm \hat{1})$ depending on the sign of J_{ij}, arrows represent the local field operator $h_i(\hat{\sigma}_i^z \pm \hat{1})$ again depending on the sign of h_i. In the example here, from top to bottom, the sites have local external fields which are negative, positive, positive, positive, and negative respectively. The signs of $J_{i,i+1}$ are positive, negative, negative, and positive from the top to bottom.

The local update procedure is easier to construct as an insertion/removal process (Fig.4.2). Namely, if $\hat{W}_{k'(l)}$ is a (non-null) operator which is *not* a transverse field operator, it is removed with probability

$$P_{\text{remove}} = \min\left[1, \ \frac{L - n + 1}{\beta(h_{\text{tot}} + J_{\text{tot}} + \Gamma_{\text{tot}})}\right], \tag{4.26}$$

and becomes a null operator $\hat{1}$. Here, $h_{\text{tot}} := 2\sum_i |h_i|$, $J_{\text{tot}} := 2\sum_{i,j}|J_{ij}|$, and $\Gamma_{\text{tot}} := \sum_i |\Gamma_i|$. The coefficient 2 comes from the adjustment with adding/subtracting $\hat{1}$ for those operators, which effectively doubles their probabilistic weights. The transverse field operator is not removed in the same way, since removing only one transverse field operator will create a configuration violating the periodic boundary condition.

On the other hand, if $\hat{W}_{k'(l)}$ is a null operator, we insert an operator (either an interaction, transeverse suppering, or a local field operator) with probability

$$P_{\text{insert}} = \frac{2|h_i| \text{ or } 2|J_{ij}| \text{ or } |\Gamma_i|}{h_{\text{tot}} + J_{\text{tot}} + \Gamma_{\text{tot}}}, \tag{4.27}$$

for each different type of operators. One should be able to see that by comparing equations (4.17), (4.26), and (4.27), the detailed balance condition Eq.(4.3) is satisfied. The insert/remove process could be done either on random places for l, or it could be swept from $l = 1$ to $l = L$. Both of them satisfy the balance condition[*10].

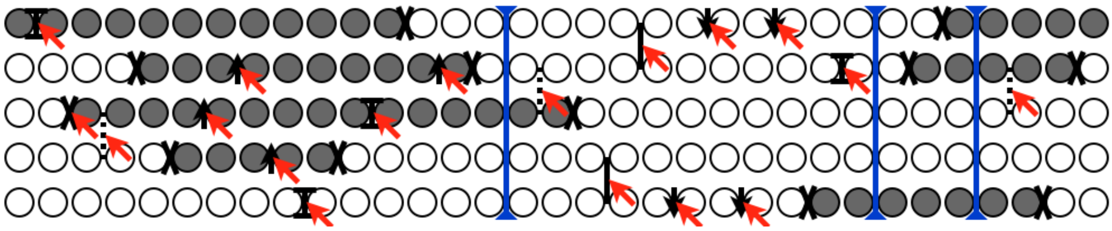

Fig. 4.2: The red arrows show which operators to consider removing in a local update. The blue lines show where we need to consider inserting an (non-null) operator. Note that all the operators except for X (transverse field operators) has a chance of getting removed, and Xs are untouched.

The second phase is the global update phase, where the *spins* in the corresponding diagram are flipped with appropriate probability. Not only the operators but also the basis $|\boldsymbol{\sigma}\rangle$ is changed, guaranteeing ergodicity of the entire Markov chain. Clusters are defined as expressed in Fig.4.3, and "local field operator free" clusters are flipped with probability 1/2. Since each transverse field operator and transverse supporting operator is generally an edge of two different clusters, they change according to the situation of the two clusters being flipped or not. If both or neither of the clusters are flipped, the operators will not change. Conversely, if only one cluster is flipped, a transverse field operator $\Gamma_i \hat{\sigma}_i$ will become a transverse supporting operator $\Gamma_i \hat{1}_i$ and vice-versa. This rule simply reflects the consistency of the operators with the spin diagram. The important point is that both operators have exactly the same probabilistic weight, and flipping the cluster with probability 1/2 achieves the equidistribution. Also, when a cluster including a spin on the boundary (of the Trotter direction), the basis state $|\boldsymbol{\sigma}\rangle$ is changed accordingly.

The two phases of local and global update compose a single Monte Carlo step. By repeating them, we obtain the equilibrium distribution of the quantum state at inverse temperature β.

4.2 Monte Carlo Techniques for Simulating Quantum Annealing

Both of the quantum Monte Carlo (QMC) techniques we introduced so far could be applied similarly to stoquastic systems. However, the techniques themselves are methods for obtaining the equilibrium state of the system, which means that it does not necessarily reflect the dynamics of the system. Since QA (and also AQC) is a framework defined by

[*10] A sequential sweep is said to *weakly* break the detailed balance condition, and is expected to be more efficient [96].

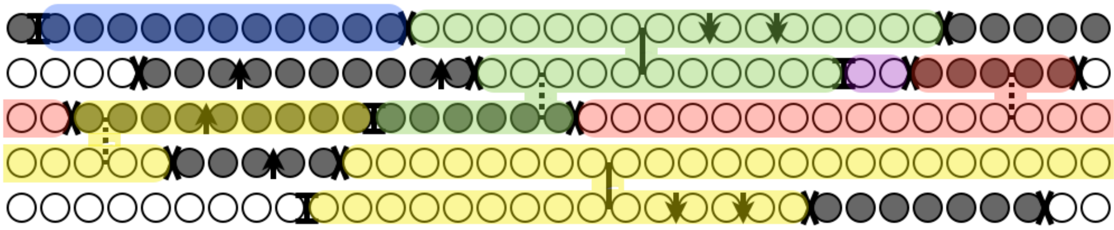

Fig. 4.3: The colored regions show examples of clusters. Clusters finish at operators I or X, and will be connected by J_{ij} interactions as in the red cluster. If the cluster contains arrows like in the green or yellow clusters, that cluster would not be flipped.

the dynamics of a quantum system, a simulation of it seems to require a dynamical approach. Naively using the the Schrödinger equation (Eq.3.1) however, will always require exponential time for computation, since the Hilbert space grows exponentially in the first place.

Here in this section, we will review several different ways of simulating QA by Monte Carlo algorithms. Although they do not have immediate connection to the dynamics of the system, we will review how they relate in the adiabatic limit.

4.2.1 Simulated Quantum Annealing

SQA is the most straightforward way of simulating QA (and thus the straightforward naming). The basic idea is simply to apply the QA procedure to QMC. More precisely, the protocol could be formalized as follows.

Definition 18. Simulated Quantum Annealing (SQA)

An SQA algorithm for a stoquastic Hamiltonian $\hat{H}(\Gamma) = \hat{\mathcal{F}}(\{\hat{\sigma}_i^z\}) + \Gamma \sum_i \hat{\sigma}_i^x$ is performed in the following way.

1. Simulate the equilibrium state of $\hat{H}(\Gamma)$ for large enough Γ and β, using QMC.
2. After certain amount of Monte Carlo steps, decrease Γ for a finite value $\Delta\Gamma$.
3. Continue this until $\Gamma = 0$, and then measure the values of each spins in the z basis.

Of course, although we used the format of $\hat{H}(\Gamma) = \hat{\mathcal{F}}(\{\hat{\sigma}_i^z\}) + \Gamma \sum_i \hat{\sigma}_i^x$, there is no problem in considering the form of $\hat{H}(\lambda) = (1-\lambda)\hat{\mathcal{F}}(\{\hat{\sigma}_i^z\}) + \lambda \sum_i \hat{\sigma}_i^x$.

One important point to mention is that SQA has a *different* dynamics compared to the actual QA dynamics obeying the Schrödinger equation [97]. Therefore, **a polynomial energy gap will *not* imply SQA in polynomial time** necessarily, although it implies polynomial time computation for the actual annealer. It is known that when using PIMC for SQA, the reducing speed of

$$\Gamma(t) \gtrsim \frac{L}{\beta}(t+2)^{\frac{-2c}{N}}, \tag{4.28}$$

will guarantee that the SQA dynamics will remain in the instantaneous equilibrium state [98]. Here, c is a constant of $O(N^0)$. We can easily see that the above bound only provides us with an exponential upper bound on the time required for SQA against general stoquastic Hamiltonians. Whether if a polynomial energy gap implies a polynomial upper bound for the corresponding SQA, remains an open problem.

Finite Temperature

Another point which should be noted is that QMC simulations in general, are simulations for *finite* temperature. One straightforward indicator is to set β larger than the energy gap $\Delta E_1 := E_1 - E_0$ [71, 164]. This is because using the energy eigenbasis, the partition function could be written as

$$Z = \sum_n e^{-\beta E_n} = e^{-\beta E_0}(1 + \sum_{n=1} e^{-\beta \Delta E_n}), \tag{4.29}$$

where $\Delta E_n := E_n - E_0$. Thus, if $\beta > \Delta E_1^{-1}$, the second term becomes negligible compared to the first term, and the equilibrium state could be seen as an approximation of the ground state. Under this setting, as long as the minimum energy gap $\Delta E_{\min}$ is at most polynomially small, setting β polynomially large is sufficient for guaranteeing that the equilibrium state is effectively the ground state.

4.2.2 Exchange Monte Carlo

Despite the similarity with the actual QA process, SQA only had a rigorous proof on its running time assuming that the Monte Carlo algorithm always reaches the equilibrium distribution close enough [98]. In general, the equilibration time of Monte Carlo algorithms becomes extremely long for “Glassy” systems, which have many local optimums in the energy landscape. The stochastic process usually gets trapped into one of the many valleys, ending up in a long equilibration time. In this thesis, later in chapter 6, we will consider an NP-hard problem with a unique solution, which also has the glassy property. Therefore, a natural strategy would be to accelerate equilibration of the Markov chain[*11]. The exchange Monte Carlo method (EMC; also known as parallel tempering, or replica exchange) [99] is one such algorithm, which is simple yet powerful, and is commonly used for accelerating equilibration of glassy systems. We will review the algorithm in this section, and explain how it is applicable to simulating QA.

The Basics

Let’s think of a classical Ising system for the sake of simplicity. At high enough temperature, Monte Carlo algorithms equilibrate fastly and thus it is easier to simulate the equilibrium state. This is because low β suppresses the role of complex energy landscapes. The idea of EMC is to use the fast equilibration at higher temperature, to achieve better equilibration for lower temperature. In order to do this, we simulate the same system with different temperature simultaneously. We call these systems *replica*, since they are all identical except for the temperature (which is the parameter of interest for this case). The most simple way to do this is to divide the temperature range $[T_{\min}, T_{\max}]$ into $R-1$ equidistributed intervals, and set R replicas with temperature corresponding to the edge of the intervals (i.e $T_r = T_{\min} + (T_{\max} - T_{\min}) \times \frac{r}{R-1}$ for the r th replica). During the simulation, whenever we find adjacent replicas with “out of place configurations”, we swap those configurations. This lets the configurations being sampled to always be “in place” with the temperature. Intuitively, replicas with high temperature search the configurations space “shallow and wide”, and once they bump into a low energy configuration, they “promote” to lower temperature by the exchange, and start to search “deep and narrow”.

Obviously, to make the above idea work, the probability and timing of the replicas being exchanged should be carefully tuned. It turns out that with the detailed balance

[*11] Notice that SQA is a Markov chain *nonuniform in time*, since the transition probability varies over time along with the annealing schedule. This becomes one obstacle for analytic studies for SQA.

condition Eq.(4.3) in mind, the exchange probability could be calculated quite easily for the general case. Let β_a and β_b two (inverse) temperature values of adjacent replicas a and b . When $\boldsymbol{\sigma}_a$ and $\boldsymbol{\sigma}_b$ are configurations of the two replicas, the equilibrium probability (concerning the entire R replicas) of having those two configurations is

$$\pi_{a,b}^{(\mathrm{eq})}(\boldsymbol{\sigma}_a, \boldsymbol{\sigma}_b) = \pi_a^{(\mathrm{eq})}(\boldsymbol{\sigma}_a)\pi_b^{(\mathrm{eq})}(\boldsymbol{\sigma}_b) = \frac{1}{Z_a Z_b} e^{-\beta_a H(\boldsymbol{\sigma}_a) - \beta_b H(\boldsymbol{\sigma}_b)}, \tag{4.30}$$

due to their independence. On the other hand, the equilibrium probability of having an *exchanged* configuration would be

$$\pi_{a,b}^{(\mathrm{eq})}(\boldsymbol{\sigma}_b, \boldsymbol{\sigma}_a) = \pi_a^{(\mathrm{eq})}(\boldsymbol{\sigma}_b)\pi_b^{(\mathrm{eq})}(\boldsymbol{\sigma}_a) = \frac{1}{Z_a Z_b} e^{-\beta_a H(\boldsymbol{\sigma}_b) - \beta_b H(\boldsymbol{\sigma}_a)}. \tag{4.31}$$

Now that we have two equilibrium probabilities, we can think of the transition probability from one to another, just as we did in section 4.1.1. Namely, if we adopt the Metropolis method (Eq.4.5)[*12], the exchange probability could be written as

$$P_{a \leftrightarrow b} = \min\left[1, \frac{\pi_a^{(\mathrm{eq})}(\boldsymbol{\sigma}_b) \cdot \pi_b^{(\mathrm{eq})}(\boldsymbol{\sigma}_a)}{\pi_a^{(\mathrm{eq})}(\boldsymbol{\sigma}_a) \cdot \pi_b^{(\mathrm{eq})}(\boldsymbol{\sigma}_b)}\right] = \min[1, e^{\Delta\beta\Delta H}], \tag{4.32}$$

with $\Delta\beta := \beta_a - \beta_b$ and $\Delta H := H(\boldsymbol{\sigma}_a) - H(\boldsymbol{\sigma}_b)$. Notice that because of the classical setting, the equilibrium probability is in a simple for $e^{-\beta H}$ which allows the final simple expression. Reconsideration will be needed when we apply this to the quantum case, but the above equation is always correct up to the second expression.

Simulating QA

The EMC method provides us a sophisticated alternative to the SQA, for simulating QA. The idea is to simply divide the annealing parameter $\lambda \in [0, 1]$ into $R - 1$ intervals, and have R replicas along the λ axis. The essence of the EMC method was to use rapid mixing regions to accelerate equilibration in the slow region, and the picture still holds for QA. Now the strong transverse field (λ large) region is rapidly mixing, and serves similar role as the high temperature region as in the classical example. All QMC methods *map* a quantum (stoquastic) system into some type of a classical system, whether if it is an $N \times L$ classical spin, or if it is a string of L operators). The classical configurations have probabilistic weight, so we can directly apply Eq.(4.32) and have exchanges along the λ axis.

This simulation with EMC, obviously does not have an immediate connection with the dynamics of QA. Rather, it is a direct simulation of the instantaneous ground state[*13], which QA should follow in the adiabatic limit. Since this is the “ideal” path which obtaining the ground state is guaranteed by the adiabatic theorem, knowing properties of the ground state provides us information about the difficulties that the QA may face, e.g. quantum phase transitions. For instance, if an evidence of an exponentially small energy gap is found, it shows that the ideal AQC will require exponential time.

It is known that the EMC method greatly accelerates equilibration for systems which have multiple basins in the energy landscape [100]. As we will see in the next chapter, hard combinatorial optimization problems are known to have rugged energy landscapes, and thus EMC methods are used for most simulations of QA [75, 101, 102].

[*12] Of course, it is possible to use the heat bath method instead. However, the heat bath method becomes more efficient when there are multiple possible configurations as the candidate. In this case, the random process is to decide whether to swap the configurations or not, and thus the Metropolis method is more advantageous.

[*13] Setting the temperature low enough allows us to sample the equilibrium state as an approximation of the ground state.

4.2.3 Population Methods

Diffusion Monte Carlo (DMC) is also a method for simulating QA, which contrary to its naming, is quite different from other QMC methods. In this section, we will briefly introduce one type of DMC [103].

Let us consider a stoquastic Hamiltonian as in Eq.(4.6)

$$\hat{H}(t) = \hat{\mathcal{F}}(\{\hat{\sigma}_i^z\}) - \Gamma(t)\sum_{i=1}^{N}\hat{\sigma}_i^x =: \hat{\mathcal{F}}(\boldsymbol{\sigma}) - \Gamma(t)\sum_{i=1}^{N}\hat{\sigma}_i^x,$$

which by definition, only has non-positive off-diagonal elements. Adding a constant term to the Hamiltonian does not make difference to the physics, therefore we can add $N\Gamma - \min_{\boldsymbol{\sigma}}[\mathcal{F}(\boldsymbol{\sigma})]$ as a constant, to ensure that $\hat{H}$ has the property of

$$\sum_{\boldsymbol{\sigma}'}\langle\boldsymbol{\sigma}'|\hat{H}|\boldsymbol{\sigma}\rangle \geq 0 \ (\forall\boldsymbol{\sigma}), \tag{4.33}$$

which becomes important soon. Now, the *imaginary time* schrödinger equation

$$\frac{d}{d\tau}|\psi(\tau)\rangle = -\hat{H}(\tau)|\psi(\tau)\rangle, \tag{4.34}$$

is simply a rewriting of Eq.(3.1) with $t = -i\tau$ corresponding to a Wick rotation. The strategy of DMC is to regard the above equation as a diffusion equation, and apply numerical approachs as *population* algorithms to simulate the diffusion equation. The infinitesimal imaginary time evolution for $\Delta\tau$ could be written as

$$\begin{aligned} e^{-\Delta\tau\hat{H}} &= \hat{1} - \Delta\tau\hat{H} + O(\Delta\tau^2) && (4.35)\\ &= \hat{1} - \Delta\tau\hat{\mathcal{F}} + \Delta\tau\Gamma\sum_i\hat{\sigma}_i^x + O(\Delta\tau^2). && (4.36)\end{aligned}$$

Because of the stoquasticity and the added constant ensuring Eq.(4.33), if we take $\Delta\tau$ sufficiently small, the matrix in Eq.(4.36) can satisfy

$$0 \leq \langle\boldsymbol{\sigma}'|e^{-\Delta\tau\hat{H}}|\boldsymbol{\sigma}\rangle \leq 1 \ \ (\forall\boldsymbol{\sigma},\boldsymbol{\sigma}') \quad \text{and} \quad 0 \leq \sum_{\sigma'}\langle\boldsymbol{\sigma}'|e^{-\Delta\tau\hat{H}}|\boldsymbol{\sigma}\rangle \leq 1 \ (\forall\boldsymbol{\sigma}'), \tag{4.37}$$

which are the conditions of being a "substochastic" matrix. It is the same as the transition matrix in Eq.(4.1), except for the fact that the probabilities do not add up to 1. To simulate this diffusion equation, we first prepare "walkers" in the z basis, each of them representing an *amplitude* of the basis state $|\boldsymbol{\sigma}\rangle$ for the stat vector $|\psi(\tau)\rangle$. Each walker in position $|\boldsymbol{\sigma}\rangle$ will have the chance of either

1. staying at the same basis state $|\boldsymbol{\sigma}\rangle$ with prbability $1 - \Delta\tau\mathcal{F}(\boldsymbol{\sigma})$,
2. flipping one qubit to become a new basis state $|\boldsymbol{\sigma}'\rangle$ with probability $\Delta\tau\Gamma$, or
3. getting erased with probability $\Delta\tau(\mathcal{F}(\boldsymbol{\sigma}) - N\Gamma)$.

By this transition probability, the population of the walkers will decrease probabilistically. Thus, normalizing the above possibility conditioned to be surviving is the actual probability used in numerical algorithms. This way, we are able to sample the *amplitudes* of the imaginary-time evolving state vector $|\psi(\tau)\rangle$.

Notice that this algorithm takes advantage of the fact that the ground state of a stoquastic Hamiltonian could be expressed by non-negative real amplitudes in the z basis, i.e. $|\Psi_0\rangle = \sum_{\boldsymbol{\sigma}} a_{\boldsymbol{\sigma}} |\boldsymbol{\sigma}\rangle$ with $\forall a_{\boldsymbol{\sigma}} \geq 0$. While this method simulates the *imaginary* time evolution and not the real time evolution, they both end up in the ground state when the change in the Hamiltonian is adiabatic. This means if the DMC is successful, it will have all the walkers in the single ground state of $\mathcal{F}$. This is how DMC simulates QA.

4.3 Examples: The Computational Complexity Perspective

So far, in this chapter, we have seen how stoquastic QA could be simulated using various techniques, such as QMC or DMC. The central question is that *if a stoquastic QA has only polynomially small energy gap (and thus requires only polynomial time), would that be simulable by these classical algorithms also in polynomial time*?

Rigorously answering this question negatively will require a separation between the class BPP and BQP, since QA is simulable in BQP. Obviously this approach is not realistic (since we cannot even prove the extremely intuitive separation of P≠NP!), and an alternative approach would be to find an example. An example problem which classical algorithms such as QMC or DBC requires exponential time while the actual QA only requires polynomial time will at least be a *supporting material* for the separation*[14].

On the other hand, answering the above question positively seems within reach, since it only requires a construction of a classical algorithm which its running time is bounded by the inverse polynomial of the energy gap for QA*[15]. An important caution to mention here is that stoquasticity (and thus mappability to classical systems) itself does *not* imply polynomial simulability. This is because the running time of QMC and DMC depend on the equilibration time, which can become exponentially long. Therefore, we need to put in mind that constructing such a Monte Carlo algorithm will require in general a proof of the equilibration time being polynomially bounded.

Currently, there are partial results on the simulability of stoquastic QA [104, 105, 78], but none of them is decisive. In this section, we will see some candidate examples for showing the separation, and address their problems. Then we consider how classical simulability may be proved, and also provide how this attempt fails. Although the results in this section may seem all negative, we will finally discuss what possibilities are left for proving either the simulability or separation between stoquastic QA (BstoqP) and BPP.

4.3.1 The Topological Obstruction

Here, we will briefly review the "topological obstruction" shown in previous work [106]. The central idea is to create ground states of a stoquastic Hamiltonian which has a non-trivial topology in the $N \times L$ configuration with the periodic boundary condition. Although multiple examples were shown in previous study [106], we will focus only on the most basic example, as it represents the idea clearly enough. For further details, see [106].

The Hamiltonian of the example is

$$\hat{H} = -\frac{1}{2ma^2}\sum_{x,y}\big(|x+a,y\rangle\langle x,y| + |x,y+a\rangle\langle x,y|\big) + h.c + \sum_{x,y} V(x,y)|x,y\rangle\langle x,y|, \quad (4.38)$$

where x and y are discretized coordinates with mesh a in the 2 dimensional plane. Note that unlike other Hamiltonians in this thesis, this example is not a spin system and we

*14 Of course, it is only a supporting material since the possibility of a smarter classical algorithm being able to simulate it in polynoimal time always remains. Then again, this is the best we could do, since separation of complexity classes is extremely difficult in general.

*15 Obviously, this is equivalent to proving BstoqP=BPP. This at least seems easier than prooving BPP=P, since it does not require any derandomization results.

use the position as the basis. $m > 0$ is the mass of the particle, and it is clear that the Hamiltonian is stoquastic. The annealing schedule is not straight-forward for this example, and we consider first changing the potential energy

$$V(x, y) = \mu(x^2 + y^2) + (x^2 + y^2)^2 - hx, \tag{4.39}$$

by tuning the parameters μ and h, and then turning the mass $m \to \infty$.

The precise annealing schedule is done as follows. First, we start from $\mu = h = 0$ and $m = 1$, which the potential energy is unimodal. The ground state is a superposition of $|x, y\rangle$ with $x^2 + y^2$ small enough. Next, μ is turned negative, creating a wine-bottled potential energy. Now the ground state is a superposition of $|x, y\rangle$ with $x^2 + y^2 \simeq r_{\min}^2$ where $r_{\min}$ is the radius of the wine-bottle. Then, h is increased from 0 to 1, which corresponds to tilting the wine-bottle. Finally, m is increased to ∞, analogous to decreasing transverse field. This lets the state $|x \sim r_{\min}, y = 0\rangle$ the ground state, with simply the lowest V.

The reason of QMC not being able to simulate this process efficiently is explained as follows. After μ is turned negative, the ground state becomes superpositions of the basis states along the circle with radius $r_{\min}$. Taking the mesh a small enough, the discretization could be ignored, and the effective Hamiltonian then becomes

$$\hat{H}_{\text{eff}} = -k\partial_\theta^2 + h' \cos\theta, \tag{4.40}$$

where we only focus on the basis states along the circle parameterized by θ. Now, the classically mapped configuration of QMC becomes a torus, since the Trotter direction also has periodic boundary condition. θ can vary along the Trotter direction, and have $n \in \mathbb{Z}$ windings when it comes back to the boundary. This *winding number* cannot change once μ is set to a large negative value, due to the barrier.

The central claim is that due to this topological number, QMC methods cannot sample the correct ground state, which is a superposition of all the topological numbers with appropriate amplitudes. In order for QMC to correctly sample among the different topological sectors, an exponential amount of time is needed for overcoming the energy barriers. QA on the other hand, can always remain in the ground state efficiently, since the energy gap of this system is confirmed to be only polynoimally small.

However, there is an unaddressed subtlety in this argument. The typical winding number n_{typ} could be estimated to be of $O(\beta)$. This is the grounds for the Monte Carlo sampling being trapped in the wrong (non-zero) topological sector. Nonetheless, the probability of $n = 0$ could be calculated and is of $O(\beta^{-1/2})$ meaning that it is only *polynomially* small. This implies that if we repeat the SQA process for polynomially many times, an $O(1)$ fraction of the trials will be confined in the trivial topological sector, ending up in the correct ground state at the end (which also is in the trivial topological sector). This "parallelized Monte Carlo" could be further turned into an even more sophisticated algorithm which is called population annealing [107]. It is currently unclear how well such algorithms will work against the topological obstructions raised along the above example, in general. The EMC method also may serve as a solution for this topological obstacle, since different replicas can hold configurations with different topological sectors. With the exchanges of replicas, the superposition of different topological sectors may be realized when there are enough replicas. Furthermore, as we discuss in the next section, it is known that the DMC method does not suffer from the topological obstructions [103]. This leads us to considering a non-topological obstruction.

4.3.2 The Non-topological Obstruction

The topological obstruction relies on the periodic boundary condition which is a feature most of QMC methods inevitably possess. As we have seen in section 4.2.3, the DMC method does not have a periodic boundary condition, since it does not take the trace. This means that the topological obstructions explained in the previous section, does not become a problem for DMC. This brings hope for a possibility that DMC enables polynomial simulation of a stoquastic QA with polynomial energy gaps in general. M. Jarret et al [103] constructed an example against this possibility.

An Example for the Failure of DMC

The example presented in [103] is a fairly simple quantum spin system with a Hamiltonian

$$\hat{H}(s) = \frac{1}{N}\left(\sum_i \hat{\sigma}_i^x + b(s)\sum_i \frac{1+\hat{\sigma}_i^z}{2}\right) - c(s)\hat{G}, \tag{4.41}$$

where the parameters $b(s)$ and $c(s)$ are tuned according to the annealing schedule. We will call the operator

$$\hat{G} = |111\ldots1\rangle\langle111\ldots1| \tag{4.42}$$

as the "golf hole potential" since it decreases the energy for one particular basis only. Here we are using the computational notation $|0\rangle, |1\rangle$ for the z basis. The annealing schedule starts from $b(0) = c(0) = 0$, where the ground state is just a uniform superposition of all the basis states, just as in a normal QA protocol. Then, $b(s)$ is increased linearly, letting the spins point in the $|1\rangle$ direction. $b(s)$ is fixed once it reaches $b(1/2)$ and then finally, $c(s)$ is increased linearly so that the "all down" state is preferred.

This simple model reveals the weakness of DMC, which is that it does not directly sample the probability, but rather samples the *amplitudes*. The Hamiltonian up until turning $\hat{G}$ on does not have any interaction, and thus the ground state could be written in the form of $|\Psi_0(s)\rangle = |\psi_0(s)\rangle^{\otimes N}$. If we carefully tune $b(1/2)$ to be $2/\tan[2\arccos(1-1/4N)]$, a straightforward calculation shows that

$$\left|\langle111\ldots1|\Psi_0(s=1/2)\rangle\right|^2 = \left(1-\frac{1}{4N}\right)^{2N}. \tag{4.43}$$

This probability tends to $1/\sqrt{e}$ as $N\to\infty$, while the ratio of the walkers in DMC with $|111\ldots1\rangle$ is only $\frac{e^{-1/4}}{e^{\sqrt{N/2}}}$ which is exponentially small. This becomes a problem in the next phase where $\hat{G}$ is turned on. Since no walker has the configuration of $|111\ldots1\rangle$ (as long as there are only polynomially many walkers), none of them are susceptible to the term $\hat{G}$, and thus correct sampling of the ground state is not achieved.

Numerically Measuring the Equilibration Time

While the above model shows an example of where the DMC method fails and QA succeeds, it does not become an obstacle when using QMC methods. We will see this by numerically measuring the time QMC requires for correctly sampling the ground state of the above example.

The relaxation time τ_{rel} and mixing time τ_{mix} of a Markov chain are two major measurements for quantifying the required time for equilibration. The mixing time is the time required for the Markov chain to get close enough to the equilibrium distribution $\boldsymbol{\pi}^{(\mathrm{eq})}$ in the worst case, i.e.,

$$\tau_{\mathrm{mix}}(\epsilon) := \min\{\ t \mid \sup_{\boldsymbol{\pi}(0)} \|P^t\boldsymbol{\pi}(0) - \boldsymbol{\pi}^{(\mathrm{eq})}\|_{\mathrm{TV}} \le \epsilon\}, \tag{4.44}$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation disitance

$$\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_{\mathrm{TV}}=\frac{1}{2}\sum_{\boldsymbol{\sigma}}|\mu_{\boldsymbol{\sigma}}-\nu_{\boldsymbol{\sigma}}|. \tag{4.45}$$

Setting $\epsilon<1/2$ is known to be sufficient, since for example

$$\tau_{\mathrm{mix}}(\epsilon)\leq\lceil\log_2\frac{1}{\epsilon}\rceil\tau_{\mathrm{mix}}(1/4), \tag{4.46}$$

could be shown for arbitrary ϵ. The relaxation time is defined as

$$\tau_{\mathrm{rel}}:=\frac{1}{1-\max|\lambda|}, \tag{4.47}$$

where the maximum is taken from all eigenvalues of the transition matrix P except for 1. The mixing time and relaxation time can bound each other as

$$(\tau_{\mathrm{rel}}-1)\log\left(\frac{1}{2\epsilon}\right)\ \leq\ \tau_{\mathrm{mix}}(\epsilon)\ \leq\ \log\left(\frac{1}{\epsilon\pi^{(\mathrm{eq})}_{\mathrm{min}}}\right)\tau_{\mathrm{rel}}, \tag{4.48}$$

therefore they have similar scaling. This is because $\pi^{(\mathrm{eq})}_{\mathrm{min}}$, the equilibrium probablity of the configuration with smallest possible probability, is at least of order $e^{-\mathrm{poly}(N)}$, which gives at most a polynomial factor. Our central concern if the simulation could be done in polynomial time or not, so this does not become a problem.

Calculating τ_{rel} explicitly is practically impossible since the transition matrix is exponentially large. However, for our model explained above, we are able to esitmate the relaxation time from the following argument. Let's consider measuring $\langle\hat{G}\rangle$ for the final Hamiltonian of Eq.(4.41). Starting from a random configuration, after t Monte Carlo steps, the probability distribution over configurations could be written as

$$P^t\left|\frac{1}{\Omega}\right\rangle, \tag{4.49}$$

where we borrow the bra-ket notation $|\frac{1}{\Omega}\rangle$ as the uniform distirbution among all possible Ω configurations. Thus, the expectation value for the measured $\langle\hat{G}\rangle$ after t Monte Carlo steps is $\langle G|P^t|\frac{1}{\Omega}\rangle$, where $\langle G|$ is the transposed vector which the element is 1 for configurations with the basis $|11\ldots1\rangle$, and 0 otherwise. From the definition of the relaxation time, we can write

$$P^t\,|1/\Omega\rangle\cong|\boldsymbol{\pi}^{(\mathrm{eq})}\rangle+e^{-\frac{t}{\tau_{\mathrm{rel}}}}\left(|1/\Omega\rangle-|\boldsymbol{\pi}^{(\mathrm{eq})}\rangle\right), \tag{4.50}$$

at $t\gg1$. The support of the equilibrium distribution $|\boldsymbol{\pi}^{(\mathrm{eq})}\rangle$ could be regarded as the configurations containing the basis $|11\ldots1\rangle$. By multiplying $\langle G|$ from the left, we obtain

$$\langle\hat{G}\rangle\cong1+e^{-\frac{t}{\tau_{\mathrm{rel}}}}\left(\langle G|1/\Omega\rangle-1\right)\cong1-e^{-\frac{t}{\tau_{\mathrm{rel}}}}, \tag{4.51}$$

since $\langle G|1/\Omega\rangle$ is exponentially small. This allows us to estimate τ_{rel} from measuring $\langle\hat{G}\rangle$ over different Monte Carlo steps. Similar arguments could hold for different quantities, as long as they capture the slowest mode (e.g. the magnetization in the case of ferromagnetic systems in the ordered phase). For the current case, $\langle\hat{G}\rangle$ surely has the slowest mode since it quantifies if the Markov chain has visited an exponentially small region in the configuration space.

Figure 4.4 shows the equilibration of $\langle \hat{G} \rangle$ for different system size N. We are only focusing on the final Hamiltonian $s = 1$ of Eq.(4.41), which means we are neither using SQA nor EMC, and only simple QMC (in this case, we used SSE instead of PIMC). We take the average over 100 samples for each system size, for every Monte Carlo step. The temprature was set as $\beta = N$, since the minimum energy gap is explicitly calculated for this system to be $\Delta E_{\min} = \frac{2}{N}$ [103], and is sufficient for observing the ground state. We

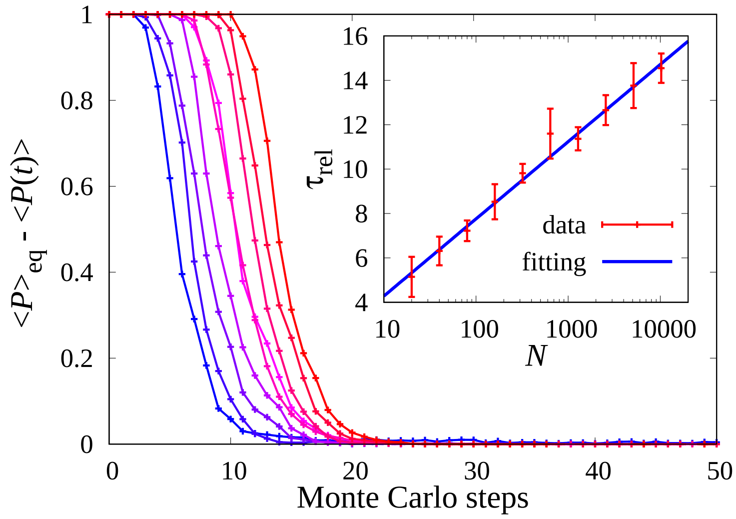

Fig. 4.4: The equilibration of $\langle \hat{G}(t) \rangle$ where t is the Monte Carlo step. From blue to red, the graph represents system sizes $N = 20, 40, 80, 160, \ldots, 10240$. The inset shows the equilibration time assuming Eq.(4.51). The errorbars are drawn using the bootstrap method.

clearly see that the Monte Carlo steps needed for equilibration only shifts additively as weincrease the system size exponentially, and the Monte Carlo steps needed forequilibration in this case only scales logarithmically. Note that 1 Monte Carlo stepmeans an updatefor theentire L operators (or similarly the update of the entire $N \times L$ classical spins for PIMC), meaning that a single Monte Carlo step already takes $O(N)$ time.

The above result has two implications. First, it shows that the simple example presented in [103], does not apply for QMC methods. This means that while DMC does not suffer from the topological obstructions but with golf-hole potentials as in this example, the situation is completely opposite for SQA. The problem of whether stoquastic QA can be polynomially simulated by classical algorithms turns into a problem of whether these two different types of obstructions could be "put together" for establishing an obstruction which both type of algorithms suffers, of if the algorithms could somehow be "put together" overcoming both types of obstacles. This point will be discussed further in the next section.

The second implication is that QMC methods do not necessarily suffer from the Arrhenius law (Eq.3.4). This could be intuitively understood as the effect of tunneling effect in quantum mechanics. While the Arrhenius law is derived from evaluating the time required for a (classical) stochastic dynamics, this does not apply for quantum systems in general. For example, the classical system which a stoquastic Hamiltonian is mapped to (Eq.4.13), has an anomalous term which has a temperature dependence $(\log \tanh \frac{\beta\Gamma}{L})/(2\beta)$. The forrmagnetic interaction along the Trotter direction becomes infinitely weak, when other parameters are fixed (or even when we increase L with fixing β/L). This prevents the Monte Carlo dynamics in the Trotter direction from freezing, avoiding the Arrhenius law.

This point is crucial if we consider proving simulability of stoquastic QA via Monte Carlo methods. When the system has a polynomially small energy gap, the inverse temperature β needs to be polynomially large for sampling the effective ground state. If the Arrhenius law Eq.(3.4) were to hold for QMC systems as well, this would immediately imply exponentially long equilibration times for such systems.

Proving Simulability: What is the Most Likely Scenario?

We have so far seen that both topological and non-topological obstructions were not powerful enough to show stoquastic advantage. More precisely, topological obstructions [106] could be avoided by DMC methods [103], and also we argued that repeating SQA, population annealing, or EMC can overcome topological obstructions. Non-topological obstructions [103], on the other hand, were not a problem at all for QMC methods, as seen in the previous section.

What implication do these examples have for the situation regarding computational advantages of stoquastic QA? So far, both of the proposed obstacles seems to be avoidable by using either repeated SQA, population annealing, or EMC. Obviously, whether if these algorithms are able to polynomially simulate stoquastic QA in general remains an open problem. Even if one algorithm among them could be proven to be able to efficiently simulate stoquastic polynomial QA, that will imply the collapse of BstoqP to BPP.

In the following, we will present some reasons for seriously considering the possibility of proving BstoqP=BPP by using EMC, and then demonstrate an example which the strategy fails.

Using EMC to prove BstoqP=BPP seems a possible strategy for several reasons. Firstly, SQA and population annealing both are not MCMC in the strict sense (the transition matrix changes along time for both algorithms), and thus has intrinsic difficulty in evaluating efficiency. EMC on the other hand, is an MCMC in the strict sense, and many results for Markov chains could be applied. Secondly, and more importantly, the notion of *exchange probability* in EMC and *fidelity* for quantum states seems to have similarities. For any of the above three algorithms, one will need to discretize the annealing process λ. Let's assume that the annealing schedule $\lambda : 1 \to 0$ is divided into $R+1$ steps from $\lambda_0 = 1$ to $\lambda_R = 1$. Since mapping to classical systems is guaranteed from stoquasticity, the remaining issue is the evaluation of the relaxation/mixing time. Intuitively, we expect that "if the ground states of adjacent parameters λ_i and λ_{i+1} are *close* enough, sampling one efficiently should also imply efficiency for the next one". This point seems to be able to made rigorous using EMC and exchange probabilities. Namely, we can prove that if the energy gap is at most polynomially small, the *fidelity susceptibility* will be at most polynomially large [*16]. Then we can further show that taking R polynomially will suffice for making the ground states of adjacent replicas arbitrarily close. With the above intuition, this result seems to suggest that with EMC with polynomially many replicas, the adjacent replicas become close enough for frequent "exchanges", allowing fast equilibration for the entire system (with all replicas).

The following example demonstrates a situation where QA could be performed in polynomial time, while the EMC method takes exponential time. This shows that the above argument is wrong, which will be discussed further in the next section. The Hamiltonian we use has the same form as Eq.(4.41) but with an extended annealing schedule. Now we divide the annealing schedule $s : 0 \to 1$ is now divided into three phases, and the first two are the same as the example before. In the third phase, after the golf-hole potential $\hat{G}$ is turned on completely, we now decrease the longitudinal field $b(s)$ to 0. While this

[*16] The calculation, including the definition of fidelity and fidelity susceptibility will be provided in chapter 6.

system is not efficiently simulable by DMC methods (since it is trivially harder than the last example), it is also evident that QMC will not be able to find the final ground state on its own. On the other hand, since the minimum energy gap is again $2/N$, QA is able to reach the final ground state in polynomial time. The expectation is that EMC will also be able to simulate the ground states throughout the process efficiently, since the sampling for $s = 2/3$ is easy.

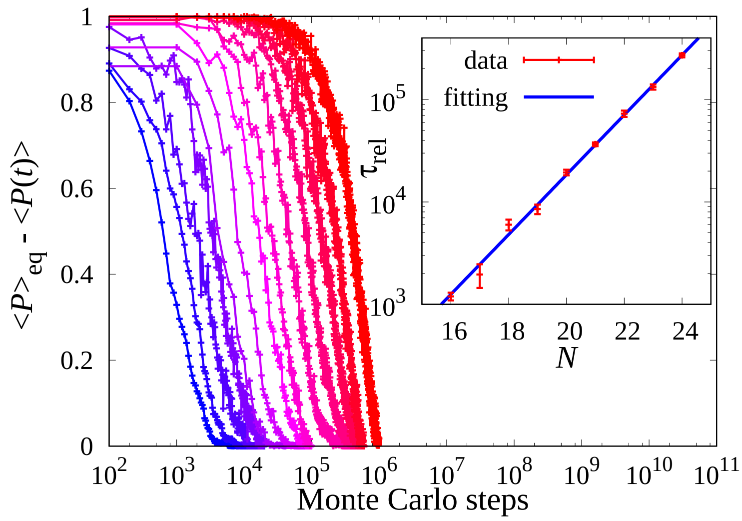

Fig. 4.5: The equilibration of $\langle \hat{G}(t) \rangle$ at $s = 1$ where t is the Monte Carlo step. From blue to red, the graph represents system sizes $N = 12, 13, 14, 15, \ldots, 24$. The inset shows the equilibration time assuming Eq.(4.51). The errorbars are drawn using the bootstrap method.

Contrary to the above argument, figure 4.5 clearly shows the exponential growth of the equilibration time for the current model. We note that this result is gained after tunings of the replicas were done. Namely, the number of replicas R, the way of dividing the interval $s \in [0, 1]$, were optimized in accordance to [108]. This immediately implies that EMC in its most straightforward form, cannot be used for proving the claim BstoqP=BPP.

The computational complexity theoretic implication will be discussed further in the next section. Here, we will raise few points which may be interesting on its own. First, this example shows the case where the equilibrium time differs greatly for different replicas despite frequent exchange. We observe that the exchange rate of the EMC method for our example is high enough ($\gtrsim$0.6), meaning that the exchange procedure is accepted more than rejections. This means that we found a concrete example which exchange does not help. In the third phase ($s > 2/3$), this model changes the longitudinal field, meaning that PIMC configurations with *fewer* spins pointing the $|1\rangle$ direction will "promote" to the large s side with exchange. This obviously does not help the final replica to find the configuration with $|11\ldots1\rangle$. It is interesting that the above explanation only works for PIMC and not necessarily for SSE, nevertheless our result shows the conclusion is same for SSE as well. SSE only deals with operators, and configurations with less *longitudinal field operators* will get exchanged to the larger s region. Here, the lack of longitudinal field operators should be playing a similar role to the lack of $|1\rangle$ spins in PIMC.

Secondly, this example may be regarded as the quantum version of the example presented in [100]. In [100], a classical spin system with a potential similar to that of the golf-hole potential $\hat{G}$ in our example was presented (called the "golf course landscape"), to explicitly see that the EMC has exponentially long equilibration time for that model.

While the model studied in [100] is trivially hard since the ground is exponentially hard to find, our model has the longitudinal field as a hint, which enables the equilibration time logarithmic in N at $s = 2/3$. The fact that a rapid mixing parameter region is well exchanged with a slow mixing region, and the slow mixing region remains slow shows a new type of weakness the EMC method has.

4.3.3 Discussions

We have seen some examples so far, which exhibited difficulties which classical Monte Carlo algorithms face for simulating stoquastic QA. The first topological obstruction provided an example of where the QMC method fails when used alone. This was not an obstruction for DMC methods, and the situation for other algorithms such as EMC or population annealing remains open (at least for the example we introduced, it can be shown that both algorithms will not have a problem). The second obstruction we saw was non-topological, designed to fail DMC. We numerically see that this example however becomes extremely easy when we use QMC methods.

Thus we have two opposite examples, one where QMC fails but DMC does not, the other vice-versa. If we can combine the two obstructions, it may result in an obstruction where *both* algorithms fail. On the other hand, if we could somehow combine the algorithms, that may result in an algorithm overcoming both obstructions. EMC was one of such candidate algorithm (and indeed overcomes both of the first two obstructions), but we presented the third example where EMC takes exponential time.

This result refutes the possibility of EMC being able to simulate all polynomial stoquastic QA efficiently. While EMC fails for our last example, SQA actually does not, and the hope of proving BstoqP=BPP currently relies on algorithms with repeated SQA or population annealing.

The fact that BstoqP=BPP cannot be proved straightforwardly in this way, together with the existence of naturally arising stoqMA-complete problems, may hint the possibility of BstoqP≠BPP. While this possibility is obviously open mathematically (even P=NP is mathematically open!), it is arguable that the situation for BstoqP=BPP is closer to that of believing P=BPP. There are derandomization results on MA=NP[32], where MA is the probabilistic analogue of NP. This means that under some plausible assumptions MA is known to equal NP while there are naturally arising MA-complete problems. Therefore, the existence of naturally arising complete problems by themselves will not serve as strong evidence for the complexity class being separated. Since even trying to show P=BPP which is already believed from derandomization argument is already extremely difficult, it is not hard to believe that BstoqP=BPP is in the same position*17.

Physically, our result may be regarded as showing the necessity of "dynamics" up to some level for simulating QA. This is because both of the remaining algorithms for showing BstoqP=BPP (population annealing and repeated SQA) are dynamical in the sense that they have time-varying dynamics. While this idea is attractive since it may lead to deeper understanding of the relationship between stochastic dynamics and statics, it remains a speculation at this point.

While the weakness of EMC we have found in this study suggests that EMC cannot be efficient for *arbitrary* stoquastic systems, we should be careful that the example we studied was artificial. It is open whether if a similar situation could naturally arise when the general form of QA Eq.(3.7) is considered. Furthermore, previous studies on classical

*17 While the P=BPP problem asks whether if randomness (at all) makes essential difference in computation, the BstoqP=BPP problem asks if stoquastic randomness could be simulated by classical randomness, which appears far more tractable intuitively.

systems [100] shows that while EMC may not be so efficient for golf-hole like energy structures, it greatly enhances equilibration in a glassy energy landscape. This Justifies our use of EMC in chapter 6, since the NP-hard problem has an energy structure close to that of glass rather than a golf-hole.

Chapter 5

The Maximum Independent Set Problem and its Classical Complexity

In this chapter, we will see how classical computational hardness appears in physical systems, by focusing on the Maximum Independent Set (MIS) problem which is NP-hard, in a purely classical setup. This will also become our starting point when we later see quantum annealing results for the same MIS problem, which would be our model for seeing how classical computational hardness appears in quantum systems. More precisely, it is necessary to know how typical-case complexity of the MIS problem is related to the statistical-mechanical property of the system for various ensembles.

We will first discuss the RS-RSB/easy-hard correspondence, a notion which is only partially confirmed up until today. This relates the phase transitions (obtained from statistical mechanics viewpoints) with the average-case complexity of NP-hard problems. We will then see that while the correspondence does not seem to hold for a rather exotic instance distribution, but will be restored when we consider a more sophisticated algorithm which we introduce. The new algorithm achieves exponential speed up for some parameter regions, resulting in polynomial time computation up to the precise RS/RSB phase transition point. Finally, we will discuss what our results suggest for the average-case complexity in statistical physics and future directions.

Our results presented in this chapter are based on the work published in [109].

5.1 The Maximum Independent Set Problem

We first introduce the MIS problem in this section.

5.1.1 Formalism and NP-completeness

In order to introduce the MIS problem, we first define what an independent set of a graph is. A graph[*1] $G = (V, E)$ is defined by two sets, namely the vertex set V and an edge set E which is a subset of $V \times V$. For now we only focus on finite graphs where V is a finite set. Furthermore, we only consider simple graphs, i.e. graphs with no multiple edges between two vertices and no self-loops. This corresponds to not allowing multiple elements or elements in the form of (v, v) in E.

Independent sets and the MIS problem could be formalized as follows.

[*1] More precisely, an undirected finite simple graph.

Definition 19. Independent set
An independent set I of a graph $G = (V, E)$ is a subset of the vertices such that no two elements are connected in graph G.
In other words, $I \subset V,\ v_1, v_2 \in I \to (v_1, v_2) \notin E$.

Definition 20. Maximum Independent Set Problem (Decision version)
Input : An undirected simple graph $G = (V, E)$ and a positive integer m.
Output : Yes if there exists an independent set I which has size larger than m.
No if there is none.

Different versions of the MIS problem exist, and we will see the relations in the following section. Although the MIS problem seems quite simple (it is basically just coloring as many vertices as possible with the constraint that no adjacent vertices are colored), it is actually NP-complete [110], which we will omit the proof.

5.1.2 Search/Decision Relations

We can easily define the search version and also the optimization version of the MIS problem.

Definition 21. Maximum Independent Set Problem (Search version)
Input : An undirected simple graph $G = (V, E)$.
Output : (One of) the Maximum independent set I of G.

Definition 22. Maximum Independent Set Problem (Optimization version)
Input : An undirected simple graph $G = (V, E)$.
Output : The size of the maximum independent set $k_{\max}$ of G.

Since both of the versions of MIS are not a decision problem, they cannot be NP-complete by definition[*2]. Clearly, the ability to solve both types of problems (search and optimization), allows one to solve the decision version as well. Either the output $k_{\max}$ itself or the size of the output I, could be compared to the inputted value k, and the decision problem is trivially solved. Thus, similarly to the reduction explained in section 2.1.2, we can say that the optimization version and search version of MIS are *at least as hard as* the decision version under *Turing reduction*[*3]. This means that both of the two versions are NP-hard.

On the other hand, if there were a polynomial time algorithm which solves the decision version of MIS, changing the input k adaptively will allow us to compute $k_{\max}$ after polynomial number of trials (thus in polynomial time). Thus, under polynomial Turing reduction, the decision version and optimization version has equivalent hardness.

Similarly, we can show that if there were a polynomial time algorithm which solves the optimization version, the search version could be computed in polynomial time as well. This is not totally apparent. The polynomial time for the search version could be

[*2] It is even hard to think of the "decision version of the optimization version" which the size of the MIS $k_{\max}$ is proposed in the input. This formalism is (believed to be) not in NP as well, since there seems to be no way of verifying that an independent set is indeed the maximum in polynomial time. This problem is in the class Σ_2^{P}, which is a generalization of NP.

[*3] This is not exactly the same as the mapping reduction we introduced in section 2.1.2, and is called Turing reduction. Now we are using the algorithms for solving the optimization/search version of MIS as a *subroutine*. The implication is the same as the mapping reduction, that proving the harder problem to be in P will immediately imply that the easier one is also in P (or FP for function problems).

constructed as follows. After computing $k_{\max}$ for the input G, we *assume* for one site on the graph v to be included in the MIS, and remove it from G along with all adjacent vertices (which cannot be included in the MIS). We then apply the optimization algorithm for this new graph G', and see if the output is $k_{\max} - 1$. If so, this means that there is at least one MIS for the original graph G which includes the vertex v, and if not, this means that the there is no MIS for G that includes v, This provides information of whether if there exists an MIS including the vertex v. The problem is now reduced to the search problem of a graph which is truly smaller than G either way. We can repeat this procedure at most N times, to fully determine the MIS, thus solving the search version of the MIS problem.

We have seen that all three versions of MIS has equivalent hardness under Turing reduction[*4]. When QA is applied for solving the MIS problem, it is natural to consider that QA is tackling the search version. Therefore, strictly speaking, it is wrong to claim it is dealing with an NP-complete problem, but in this thesis, with the above equivalence in mind, we do not worry too much if a problem is NP-hard or NP-complete. Either classes should be hard, if we assume the NP hardness assumption.

Statistical Mechanics Formulation

All of the three versions of the MIS problem could be rewritten in the language of statistical mechanics as we explained in section 3.2. In this case of the MIS problem, the Hamiltonian (cost function) could be written as

$$H(\boldsymbol{x}) = -\sum_{v} x_v + \alpha \sum_{(v,w)\in E} x_v x_w, \tag{5.1}$$

where the variables $\boldsymbol{x} = (x_1, \ldots, x_N)$ take values in $\{0, 1\}$ with 1 meaning being included in the MIS and 0 not. α is a constant larger than 1. The precise value of α does not matter as long as we are only interested in the $T \to 0$ limit, i.e., the ground state/optimal solution. The results from statistical mechanics analysis uses eq. (6.2.1) as a starting point. Note that the Hamiltonian above could be rewritten in the form

$$H(\boldsymbol{x}) = H'(\boldsymbol{\sigma}) = \frac{\alpha}{4} \sum_{(v,w)\in E} \sigma_i \sigma_j - \sum_{i} \frac{2 - \alpha d_v}{4} \sigma_v, \tag{5.2}$$

after the rewriting the variables $x_v = (\sigma_v + 1)/2$ in Ising degrees of freedom. Here, d_v denotes the *degree* of vertex v, the number of adjacent vertices. This system only has antiferromagnetic interactions with the same stregth, and thus could be interpreted as an antiferromagnetic Ising model with local fields. The decision version of MIS corresponds to asking if the ground state energy is lower than some value E_{input}, and the optimization version corresponds to asking directly the ground state energy. The search version is obviously equivalent to searching the ground state configuration itself.

5.2 Overview: Algorithms and Phase Transitions

As we mentioned in the previous section, the MIS problem is NP-complete and thus has no known polynomial time algorithm. However, as we discussed in section 2.2, some simple algorithms actually run typically in polynomial time for easy ensembles of instances. The

[*4] This reflects the computational complexity result P=NP⇒P=PH, which is called “the collapse of the polynomial hierarchy”, where PH is the class *polynomial hierarchy*, a generalization of NP including Σ_2^{P}.

field of statistical physics dealing with mean-field spin glasses could be applied for average case complexity of optimization problems [111, 112] in order to obtain these types of results.

It was first discovered by R. Monasson et al.[113] that optimization problems exhibit phase transitions, resembling physical systems. They found that random SAT in the large size limit (i.e. $N \to \infty$ the thermodynamic limit), have probability 1 for the answer being YES in the "low constraint density" region, which suddenly drops to 0 in the "high constraint density" region. While this phenomena is now known as the "SAT-UNSAT" phase transition, statistical physics further provided a picture of a phase transition called replica symmetry breaking (RSB), which seems even more essential [114, 115, 116].

5.2.1 Replica Symmetry Breaking

Here, we will breifly review RSB. Let us consider the SK model defined in sectoin 2.2.2, which was the first spin-glass model explicitly studied [117]. The Hamiltonian

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j, \tag{5.3}$$

has quenched disorder J_{ij} as mentioned before, which makes the *order parameter* non-trivial. For the SK model, the interaction strength is drawn from a Gaussian distribution

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi J^2}} \exp\left\{-\frac{N}{2J^2}J_{ij}^2\right\}, \tag{5.4}$$

where the coefficients are for normalization and additivity of physical quantities (such as average energy). At low temperature, the SK model exhibits a phase transition where each spin freezes in a seemingly random direction. The direction which the spins freeze obviously differs from sample to sample, and may even differ completely (no negative/positive correlation) for the same sample with different initial configuration![*5] In order to capture the freezing of the spins in this case, the overlap parameter (also called the spin-glass order parameter)

$$q := \frac{1}{N}\sum_{i=1}^{N} \sigma_i^{(a)}\sigma_i^{(b)}, \tag{5.5}$$

is defined. Here, we prepare two exactly same system a and b (with identical quenched disroder $\{J_{ij}\}$). $\sigma_i^{(a)}$ and $\sigma_i^{(b)}$ represent the spin value of the i th spin for system a and b respectively. Thus, q quantifies how similar two configurations $\boldsymbol{\sigma}^{(a)}$ and $\boldsymbol{\sigma}^{(b)}$ are, taking the value between $[-1, 1]$. Since in thermal equilibrium, configurations appear according to the Boltzmann distribution eq.(3.3), q becomes a random variable. RSB is defined as a phase where the distribution of q, denoted as $P(q)$, has a non-trivial structure. Let's see this through the classical Ising model as an example. In the paramagnetic phase, the value of q is always close to 0 because of thermal fluctuation. The distribution $P(q)$ is concentrated at the center, and becomes $\delta(0)$ in the thermodynamic limit. In the low temeprature phase where the magnetization

$$m := \frac{1}{N}\sum_i \sigma_i \tag{5.6}$$

[*5] Here, we are implicitly assuming some kind of dynamics. Any valid Monte Carlo dynamics with only local updates would have the same property.

typically takes a non-zero value $\pm m_{\rm typ}$, $P(q)$ becomes double peaked at two values $\pm q_{\rm typ}$. While the expectation value of m for a finite system is always 0 due to the $\mathbb{Z}_2$ symmetry, the distribution is double peaked at two values $\pm m_{\rm typ}$ representing spontaneous magntization. Thus, if two random configurations are drawn from the Boltzmann distribution, we will have two configurations with the same magnetization with probability 1/2, or configurations with exactly the opposite magnetization with probability 1/2. We can say that $P(q)$ captures the spontaneous symmetry breaking for the classical Ising model.

If we apply an external field $h\sum_i \sigma_i$ to the Ising model, the ferromagnetic transition vanishies, and q will be concentrated in a non-zero value. This is an artifact of having a non-zero magnetization, and we could subtract $m_{\rm typ}^2$ from q, obtaining the "non-trivial" overlap. The Edwards-Anderson order paramter [118] is defined as

$$q_{\rm EA} := \max_{P(q)>0:N\to\infty} q, \tag{5.7}$$

which is equal to

$$q_{\rm EA} = \frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle^2, \tag{5.8}$$

if we restrict the distribution for taking the average $\langle\cdot\rangle$ to one basin*6. Now we can define RSB as a phase transition where $q_{\rm EA}$ becomes non-zero (after subtracting the trivial value), and there is no other order parameter*7. For the ferromagnetic transition, obviously we can capture this transition via the histogram of m, or even by $\langle m^2\rangle$, and thus is not an RSB.

The Spin-glass Susceptibility

Since the average magnetization $\langle m\rangle$ is always 0 for a finite ferromagnetic Ising model, it may be convenient to measure $\langle m^2\rangle$ for numerical simulations. This procedure is essentially measuring the magnetic susceptibility

$$\chi_{\rm l} := \left.\frac{\partial m_{\rm typ}}{\partial h}\right|_{h=0} = \frac{\beta}{N}\left(\langle M^2\rangle - \langle M\rangle^2\right), \tag{5.9}$$

where $M=\sum_i\sigma_i$, and h is the strength of the uniform magnetic field $-h\sum_i\sigma_i$ added to the Hamiltonian. Note that for finite systems, the second term $\langle M\rangle^2$ is 0. If $\chi_{\rm l}$ has a non-zero value (or equivalently if $N\chi_{\rm l}=\partial M_{\rm typ}/\partial h$ diverges in the thermodynamic limit), this means that the magnetization is nonzero, or equivalently, that the system is susceptible against uniform magnetic field.

We can consider an analogous susceptibility for spin glass, which would provide us a clearer indication of the spin-glass phase transition. By further expanding the magnetization $m_{\rm typ}$ in terms of h as

$$m_{\rm typ} = \chi_{\rm l} h + \chi_{\rm nl} h^3 + O(h^5), \tag{5.10}$$

*6 This is parallel as defining the magnetization as in eq.(5.6), and claiming that the ferromagnetic phase is where the magnetization $\langle m\rangle$ has non-zero value. For finite size systems, this is always 0 due to $\mathbb{Z}_2$ symmetry. However, in the thermodynamic limit, the system is confined in one side of the configuration space, since the energy barrier between grows as $O(N)$, resulting in exponential time to reach the other. We denote this magnetization as $m_{\rm typ}$ here. We can assume the same for $q_{\rm EA}$, where the system is confined in one of the basins.

*7 This definition becomes inaccurate when we consider systems which are not mean-field. For instance, the Edwards-Anderson model, which is the 3 dimensional version of the SK model, seems to have a non-trivial $P(q)$ but not a standard RSB [119]. In this thesis, we will only deal with mean-field models.

we can define the nonlinear susceptibility χ_{nl}, which could be measured in experiments[120]. It is simple to calculate that

$$3!\chi_{\mathrm{nl}} := \left.\frac{\partial^3 m}{\partial h^3}\right|_{h=0} = \frac{\beta^3}{N}\left(\langle M^4\rangle - 4\langle M^3\rangle\langle M\rangle - 3\langle M^2\rangle^2 + 12\langle M^2\rangle\langle M\rangle^2 - 6\langle M\rangle^4\right) \tag{5.11}$$

$$= \frac{\beta^3}{N}\left(\langle M^4\rangle - 3\langle M^2\rangle^2\right), \tag{5.12}$$

where the last equation holds when $\langle M\rangle = 0$, e.g. when there is $\mathbb{Z}_2$ symmetry. χ_{nl} diverges in the RSB phase, and thus sometimes the term nonlinear susceptibility and "spin-glass susceptibility" is sometimes used interchangeably. However, we can state this in a more precise way.

The linear susceptibility represented fragility against weak uniform perturbations in opposite directions. Now that in the spin-glass phase spins freeze in "random" directions, the perturbation should be towards random directions, which could be written as a term $\epsilon\sum_i s_i\sigma_i$ where $s_i \in \{\pm 1\}$ forms some arbitrary configuration $\boldsymbol{s}$. Then, the spin-glass susceptibility should represent the fragility against two different perturbations, and defining it as

$$\chi_{\mathrm{SG}} := \lim_{\epsilon\to 0}\frac{1}{\epsilon N}\sum_i \mathbb{E}_{\boldsymbol{s},\boldsymbol{s}'}\left[\left(\langle\sigma_i\rangle_{\epsilon,\boldsymbol{s}} - \langle\sigma_i\rangle_{\epsilon,\boldsymbol{s}'}\right)^2\right], \tag{5.13}$$

is one reasonable way. Here, $\langle\cdot\rangle_{\epsilon,\boldsymbol{s}}$ denotes the expectation value under the Boltzmann distribution of Hamiltonian $H - \epsilon\sum_i s_i\sigma_i$ and the expectation $\mathbb{E}_{\boldsymbol{s}\boldsymbol{s}'}[\cdots]$ is taken over all possible configurations with Boltzmann weight of the configurations $\boldsymbol{s}$ and $\boldsymbol{s}'$, i.e., $e^{-\beta H(\boldsymbol{s})}/Z$. The definition of spin-glass susceptibility above could be related to the nonlinear susceptibility under some conditions*8,

$$\chi_{\mathrm{nl}} = \beta^3\left(\chi_{\mathrm{SG}} - \frac{2}{3}\right). \tag{5.14}$$

This means χ_{nl} and χ_{SG} are essentially the same, and both quantities have same divergences etc. at the phase transition point. Furthermore, by expnading $\langle\sigma_i\rangle_{\epsilon,\boldsymbol{s}}$ up to first order of ϵ, we can show

$$\chi_{\mathrm{SG}} = \frac{1}{N}\sum_{i,j}\left(\left.\frac{\partial\langle\sigma_i\rangle}{\partial h_j}\right|_{h_j=0}\right)^2 = \frac{\beta^2}{N}\sum_{i,j}\left[\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle\right]^2, \tag{5.15}$$

reducing the calculation to summing the equilibrium correlation. This finally leads us to the expression

$$\chi_{\mathrm{SG}} = N \times \mathrm{Var}[q] = N\left(\langle q^2\rangle - \langle q\rangle^2\right). \tag{5.16}$$

This means that if there is a non-trivial structure in the overlap distribution $P(q)$, the spin-glass susceptibility χ_{SG} will detect it.

5.2.2 RS-RSB/Easy-Hard Correspondence

The RSB transition introduced in the last section could be found ubiquitously in combinatorial optimization problems [112]. While optimization problems could be seen as the

*8 The condition for eq. (5.14) to hold is the symmetry of $P(J_{ij}) = P(-J_{ij})$ and the absence of local fields [121]. Recently, it was shown that it also holds for systems with symmetric random local fields $\tilde{P}(h_i) = \tilde{P}(-h_i)$ [122].

task of finding the ground state, and thus corresponds to the $T = 0$ or $\beta \to \infty$ limit, we can extend problems to finite temperature and locate an RSB phase transition at some $T_c > 0$. The RSB phase transition also occurs for many problems if we change the parameter of the instance ensemble. For instance, we can think of a 3-SAT ensemble where there are N variables and M clauses. If we fix the ratio of clauses $\alpha := M/N$, and take the thermodynamic limit $N \to \infty$, the resulting ground state undergoes an RSB transition at $\alpha_c \sim 3.86$ [116]. For the MIS problems we consider, as we will see soon, the statistical mechanics formulation requires local fields, and thus we have a finite magnetization m for all finite temperatures. Therefore, we use the term “replica symmetric (RS) phase” instead of the paramagnetic phase.

The RS phase has a smooth and connected solution space, while the RSB phase has rugged non-ergodic solution space. In general, the RSB phase is realized in the parameter region with more constraints, and thus is thought to be a “harder region” as an optimization problem. We will call this perspective as the *RS-RSB/easy-hard correspondence*.

The RS-RSB/easy-hard correspondence has been shown rigorously for few specific models [124, 125]. More precisely, it is known that certain algorithms takes typically only polynomial time to run in the RS phase, and requires exponentially long time in the RSB phase. However, this does not seem to hold in general. For example, the situation is still open for the NP-complete 3-SAT. For simple algorithms such as SA, it is known that they only work up to the RSB transition point $\alpha_c \sim 3.86$, However, more complex algorithms such as “survey propagation” is known to work in polynomial time up to another point $\alpha_{\text{freeze}} \sim 4.254$ [116]. Furthermore, it has recently been shown that this new freezing phase transition α_{freeze} could be overcome as well by an even more sophisticated algorithm [123]. Since it is known that another RSB transition occurs at some $\alpha_{\text{RSB}} > \alpha_c$ for 3-SAT, the situation regarding RS/RSB and the easy/hard region remains unclear in general. Another example is known as the XORSAT problem [126, 127], which we will not explain in detail but is similar to SAT. XORSAT is known to have an RSB transition but the decision version always has a polynomitl time algorithm[*9]. This seems to clearly violate the correspondence, but possibility of restoring the correspondence via further analysis taking into account of the algorithm being used. On the other hand, some natural algorithms may fail even in the RS phase, as we will explain in the following sections. Therefore, a suited algorithm seems necessary for establishing

Therefore, a good algorithm and a suited statistical physics analysis is needed to explore the general validity of the correspondence. Our main motivation in this chapter is to seek for the “correct” algorithm exhibiting the correspondence for the MIS problem.

5.3 Algorithms for the Maximum Independent Set Problem

Since the MIS problem is NP-complete, we do not have a polynomial time algorithm for solving it. However, some simple algorithms run typically in polynomial time for easy ensembles of instances, and here we present them.

5.3.1 The Leaf Removal Algorithm

We first introduce the leaf removal (LR) algorithm for the MIS problem.

[*9] The decision version is easy because XORSAT could be seen as a system of linear equations, and deciding if a solution exists is in P by using Gaussian elimination. The search and optimization versions are also easy to answer if the decision version is YES from the Turing reductions we have seen, but no clue is gained in general if the answer is NO. This may suggest that search/optimizations are better suited frameworks for the RS-RSB/easy-hard correspondence.

In general, there may be multiple MISs for a given graph G, and we take advantage of the fact that we are only required to find *one* of them, in either the search version or the optimization version. The central idea for LR is that if there is a vertex v in G with only one adjacent vertex w, there is at least *one* MIS which contains v. This could be explained as follows. For any MIS configuration I, exactly one of the vertices v, w should be included and the other should not. Having both vertices is trivially impossible for their adjacency. An MIS configuration with neither of the vertices is also a contradiction, since we can construct a larger independent set by simply adding v in it. Now, let us consider an MIS configuration where the vertex w is included and v is not. The existence of such an MIS will imply another MIS configuration with v included and w not. This is because the size of the independent set does not change with this exchange, and no constraints (edges) will be violated since v is only adjacent to w. Therefore, there will always be an MIS configuration containing v.

The LR algorithm simply finds a vertex v with only one adjacent vertex (degree 1), and label that vertex as being "`included`" (vertices with no adjacent vertex (degree 0) will trivially be labeled as "`included`" as well.). It could be argued that this algorithm focuses only on MIS configurations which include v. Then, the one adjacent vertex of v, will be labeled "`excluded`". Now, these two vertices are *removed* from the graph G, which means vertices originally having either or both of vertices v and w as adjacent vertices will have their degree decreased by 1 or 2. This will create additional vertices with degree 1 or 0.

The algorithm continues the above procedure until there are no longer any vertices with degree less than 2. Fig. 5.1 shows a schematic diagram for this algorithm.

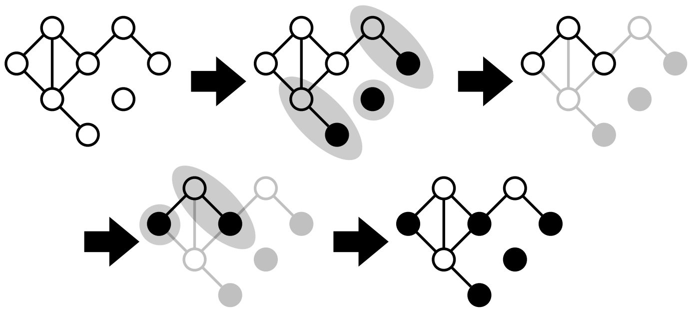

Fig. 5.1: A schematic diagram for the leaf removal algorithm. Black colored vertices are the ones decided to be `included` in the independent set. The shadowed vertices are the parts considered as the "leaves", and are removed from the graph, becoming pale. ©2017 The Physical Society of Japan (J. Phys. Soc. Jpn. **86**, 073001.)

The LR algorithm could be understood as removing the *leaves* from the graph, that do not contribute to the essential hardness of the MIS problem. By removing these leaves, new leaves may arise due to the reduction of edges. If the entire graph turns out to become a leaf, the LR algorithm successfully finds the MIS. The remainder of the graph, left after the LR algorithm stops, is called the LR-core. Note that the LR-core is well-defined independently of the order of the leaves being removed. Even if an LR-core remains, it would not be a problem as long as its size is $O(\log(N))$ or smaller, since a simple brute force search will completely determine the MIS in polynomial time. On the other hand, if the LR-core has size $O(N)$, the brute force search will take exponential time, which should be fair to say that the LR algorithm failed. In section 5.4.1 we will see that this simple algorithm may work surprisingly well for some instance ensembles.

5.3.2 The DPLR Algorithm

The LR algorithm had an obvious drawback where it was hopeless against graphs with large LR cores. It is however possible that some LR-cores may actually be "fragile", in the sense that if we perturb the LR-core by removing some vertex from it, the remaining graph becomes manageable by the LR algorithm. If this were the case, the LR-core is no longer an intrinsic obstacle. While the LR-core may appear once more, we can continually examine if the remaining LR-core is fragile or not. The idea of the Dynamic Programming Leaf Removal (DPLR) algorithm (Fig. 5.2) which we introduce here, is to combine the idea of dynamic programming (DP) with the LR algorithm in order to fully attack the fragilitiy of the LR-core. Its name is also a pun for the famous Davis-Putnam-Loveland-Logeman (DPLL) algorithm [128], which is an analogous algorithm for SAT. Dynamic programming is a common technique in computer science [137] which the main idea is to efficiently execute a "divide-and-conquer" algorithm by appropriately recording the progressive results. The procedure could be described by a recursive process, and therefore resembles to transfer matrix methods and belief propagation methods.

There is no difference between the DPLR algorithm and the LR algorithm, if the graph does not contain an LR-core. While the LR algorithm halts when it hits an LR-core, DPLR chooses one of the vertices with largest degree [138], and essentially undergoes a binary search of whether to `include` or `exclude` the selected vertex in the MIS. After fixing the label of the chosen vertex, the DPLR algorithm starts the LR protocol again. Each time reaching the LR-core, the above branching is repeated, until all the vertices are labelled either way, and the size of the independent set (not necessarily maximum) is calculated. The algorithm then goes back to the unsearched branches, constantly recording the largest-so-far independent set configuration.

The DPLR algorithm could be seen as essentially carrying out the perturbation we argued above, until the entire graph is completely broken down into leaves. Thus the DPLR algorithm becomes a general protocol which reveals the fragility of a given graph's LR-core. We will also use this algorithm for probing the hardness of a particular ensemble in chapter 6. In the following, we will see how the DPLR algorithm works typically against random instances, and probe the differences with simple LR.

5.4 Results on Random Ensembles

The random instances of the MIS problem are simply randomly generated graphs. This is one reason why we chose the MIS problem for probing the RS-RSB/easy-hard correspondence; because it has one of the most accessible input ensembles[*10]. We will first examine the Erdös-Rényi random graph, which is one of the most simple random ensembles for graphs. We will see that the RS-RSB/easy-hard correspondence holds good in this case, even by only using the LR algorithm. Then, we will go to a more complex ensemble of scale free networks, which the correspondence no longer simply holds. We will see how it is restored by using DPLR, and finally discuss the implications.

[*10] For example, a random 3-SAT instance could be seen as a random 3-uniform *hypergraph* with additional 3 degrees of freedom for each edge (similar to a directed graph). This is far more complex than a simple graph as an input.

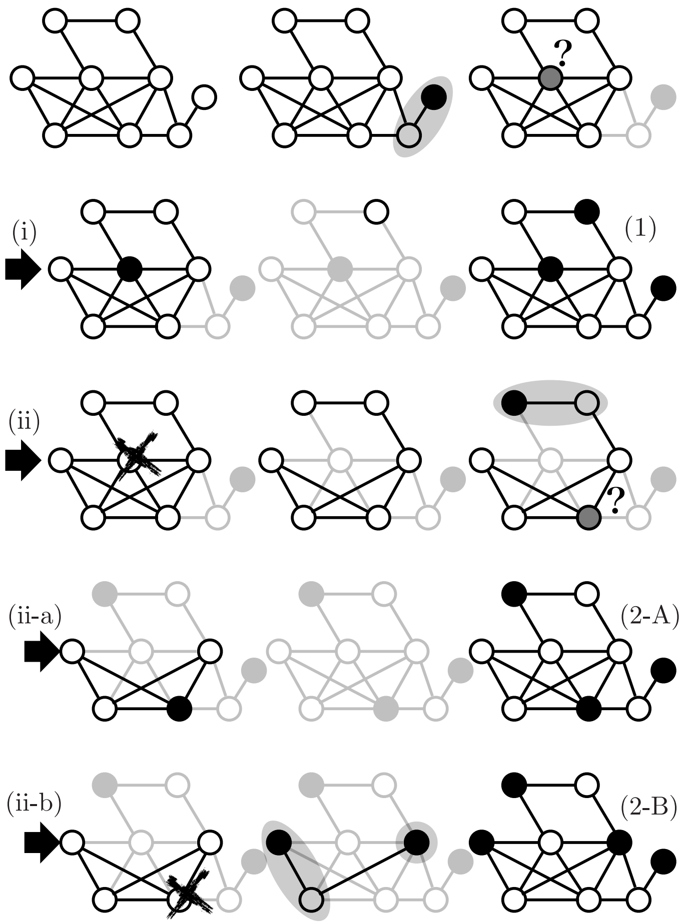

Fig. 5.2: A schematic diagram for the dynamic programming leaf removal algorithm. The gray-colored vertices are where the branching occurs, represented by the arrows. The cross mark means that the vertex is `excluded` from the independent set. The configuration and the size of the (so-far) maximum independent set found is always recorded, and is replaced whenever a larger independent set is found. In this example, there are two branches (i,ii and a,b), resulting in three different independent set configurations (1, 2-A, and 2-B). The second one (2-A) has the same size as the first one (1), so a replacement will not occur until the third configuration (2-B) is found, which has the largest size.

5.4.1 Erdös-Rényi Random Graphs

Erdös-Rényi random graphs of size N are defined as graphs with N vertices with every pair of vertices (v, w) having uniform probability p of having an edge. It could also be defined as the uniform distribution among all graphs with N vertices and $M = N^2p/2$ edges. These two ensembles match in the thermodynamic limit $N \to \infty$, just as the canonical ensemble and microcanonical ensemble.

The Erdös-Rényi random graph is known to show several phase transitions when the average degree $c = pN$ is varied. When $c < 1$, the graph is typically not connected, meaning that all connected clusters have size with only $O(\log(N))$. On the other hand, when $c > 1$ the graph has a *giant component* meaning that $O(N)$ vertices are connected as one cluster. This phase transition is known to cause the failure of a naive dynamic programming algorithm [139], while it does not affect the LR algorithm. There is a more interesting phase transition at the Napier's constant $c = e \simeq 2.718$ [129]. This phase transition has two aspects to be considered. From the physical perspective, it is an RS-to-full RSB phase transition in terms of the solution of the MIS problem*11 [130]. This means that the uniform distribution among all MISs becomes to have a nontrivial $P(q)$ and finite χ_{SG}. In the algorithmic sense, this phase transition is where the LR algorithm [131] ceases to succeed, leaving an undecided LR-core of size $O(N)$ [132]. We will write this as $c_{\mathrm{RSB}} = c_{\mathrm{LR}} = e$. Furthermore, it is shown that linear relaxation, which is another algorithm, fails at this point as well [133]. The fact that two algorithms with very different strategy start to fail at the RS/RSB transition point can be seen as a concrete example of the RS-RSB/easy-hard correspondence*12.

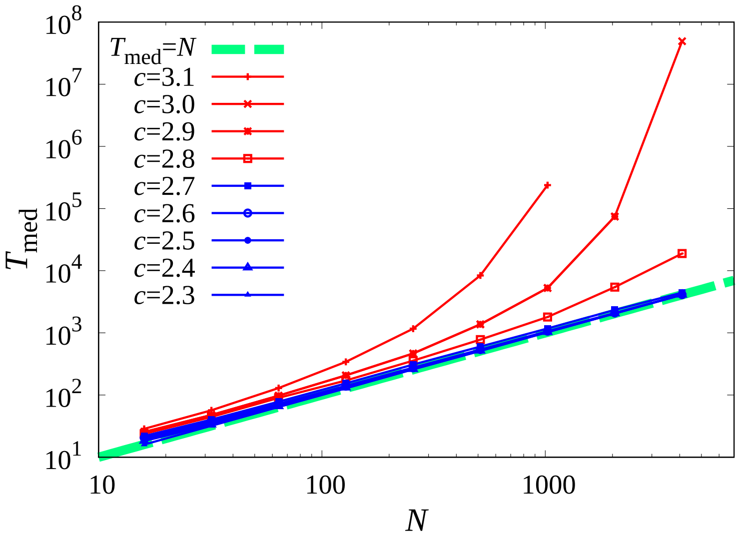

Fig. 5.3: The median running time T_{med} of DPLR on 1024 Erdös-Rényi random graphs with different sizes and different average degree c.

Numerical Results

By applying the DPLR algorithm, we can further verify this scenario of the RSB causing average-case hardness. A priori, it is possible that the LR-cores in the RSB phase is actually fragile, and DPLR allows us to compute the MIS in polynomial time. Fig. 5.3 shows that it is not the case. We plot the median running time T_{med} of the DPLR algorithm on Erdös-Rényi random graph for 1024 samples. The running time T is defined by the number of total decisions made for vertices to be `included` or `excluded`. The reason for observing the median time instead of the mean time should be obvious from arguments done in section 2.2.1. While it is difficult to claim about typical behaviours

*11 Full RSB is one of the most commonly occuring scenario of RSB. This is explained as when $P(q)$ has a continuous support between two δ peaks. The equilibrium state in full RSB has an infinite *ultrametric* hierarchical structure, which we will not go into the details. Here, it suffices to know that full RSB is the same type of RSB occuring at α_{RSB} for 3-SAT, an RSB of a type even *more* complex than the one at α_c (1RSB), which was surmountable by sophisticated algorithms.

*12 It is also interesting that the failure of linear relaxation is quite similar to that of LR, in the way that it gives back a "core" of $O(N)$ undecided vertices. However, these two cores do not match.

from the mean running time, if the median time scales exponentially, that will serve as an evidence for typical hardness.

We can see that all graphs for $c < e$ collapse into the common line $T_{\mathrm{med}} = N$. This is the trivial lower bound, and is consistent with the fact that at $c < e$, LR algorithm suffices typically, and no branching is necessary. On the other hand, we observe that all graph lines for $c > e$ have convex curves, implying super-polynomial growth of T_{med}. This means that the LR-cores are fairly robust inside all of the RSB region, requiring super-polynomial numbers of branching for completely breaking down.

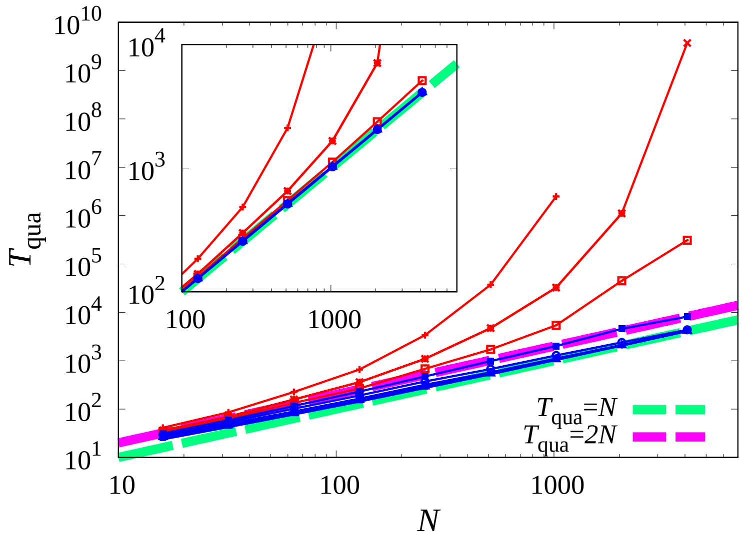

Fig. 5.4: The first quartile running time T_{qua} of DPLR on 1024 Erdös-Rényi random graphs with different sizes and different average degree c. Lines are expressed as the same as Fig. 5.3. The inset shows the third quartile running time similarly.

We further see the size dependence of the first and third quartile of the running time T_{qua} for different c in Fig. 5.4. The fact that the graphs have qualitatively similar behavior as in Fig. 5.3 suggests that the median time T_{med} sufficiently captures the average-case complexity, without explicitly calculating accordingly to the definition introduced in section 2.2.1. We will refer to the only quantitative difference seen in the first quartile of $c = 2.7$ where $T_{\mathrm{qua}} = 2N$ later. For now, we remark that a linear-to-exponential transition in typical computation time at c_{RSB} was observed for the Erdös-Rényi random graph ensemble, assuring the RS-RSB/easy-hard correspondence.

5.4.2 Scale Free Networks

The situation differs when we consider more complicated graph ensembles. The degree distribution of Erdös-Rényi random graphs is Poisson, reflecting the fact that all vertices are equal and independent. We can deviate from the Erdös-Rényi ensemble by considering other degree distributions, such as power-laws. Scale free networks are such random graphs where the degree distribution obeys power law. The Barabási-Albert model [135] provides a natural protocol which produces scale free networks. However, these scale free networks have degree correlation which makes replica analysis difficult. We can instead focus on a random graph ensemble which has the exact same degree distribution but no degree correlation. Numerically producing the ensemble is possible by following the procedure known as the configuration model [134]. From now, we will focus on such graph ensembles with the degree distribution is

$$p_k = \begin{cases} 0 & (k < m) \\ \frac{2(1-p)}{m+2} & (k = m) \\ \frac{2m(m+1)}{k(k+1)(k+2)}(1-p) + \frac{2(m+1)(m+2)}{k(k+1)(k+2)}p & (k > m), \end{cases} \tag{5.17}$$

where c is the average degree, $m := \lfloor c/2 \rfloor$, and $p := c/2 - m$. This could be seen as a configuration model of the (linearly combined) Barabási-Albert model, and thus we will call this ensemble the CBA random graph model. The absence of degree correlation allows statistical-mechanics analysis, and it is known that an RS to full-RSB transition occurs at $c_{\mathrm{RSB}} \simeq 5.239$ [136]. On the other hand, in the parameter region $c \geq c_{\mathrm{LR}} = 4$, all the vertices have degree ≥ 2, which makes the entire graph an LR-core. Thus, for this graph ensemble, the easy/hard transition for the LR algorithm disagrees with the RS/RSB transition, i.e., $c_{\mathrm{LR}} \lnsim c_{\mathrm{RSB}}$. As we will see in the following, this disagreement comes from the weakness of the LR algorithm to illustrate the correspondence for CBA random graphs. We will see in the following that DPLR achieves the correspondence.

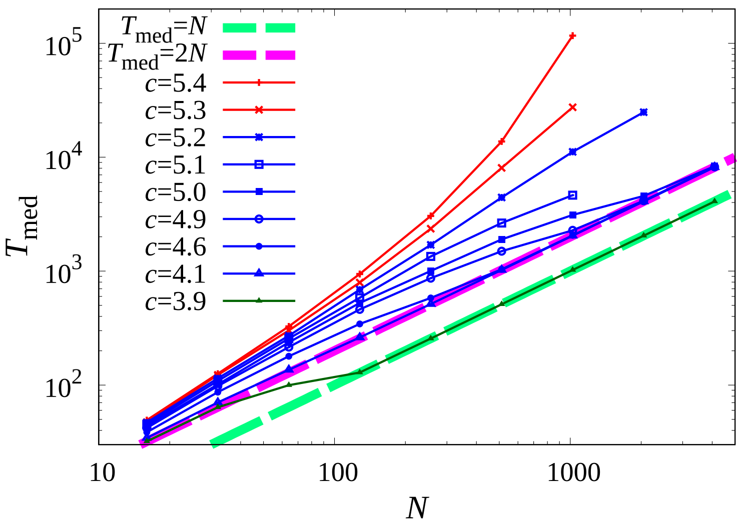

Fig. 5.5: The median running time T_{med} of DPLR on 1024 scale-free CBA random graphs with different sizes and different average degree c.

We can see from Fig. 5.5 that DPLR algorithm behaves quite differently from LR for scale free networks. Figure 5.5 shows the median running time T_{med} of the DPLR algorithm on the CBA random graphs for 1024 samples. One remarkable point about this figure for the CBA model is that graphs for $4 < c < c_{\mathrm{RSB}} \simeq 5.239$, T_{med} shows a linear growth. This means that DPLR reduces typical computation time from exponential to linear in N, which is an exponential speed up, in this parameter region when compared with naive LR. In this region, T_{med} scales as $2N$ asymptotically, with finite size effects for small system sizes, letting T_{med} larger than $2N$. This could be understood as the effect of "hub" vertices in scale free networks which have outstandingly high degrees. These hubs make the LR-core fragile to DPLR-type perturbations, since they have many neighboring vertices that will be affected when deciding the hub to be `included`/`excluded` in the independent set. When the system size is small, there will be fewer hubs, which makes the graphs robust against DPLR.

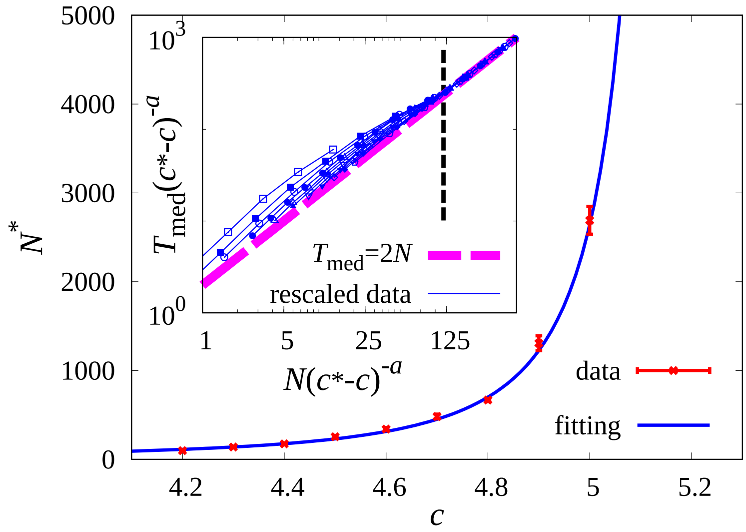

Fig. 5.6: The c dependence of the size N^* where the finite size effect becomes small enough, i.e. when $T_{\text{med}}/2N \leq 1.05$. The error bars were evaluated by bootstrap. The solid line indicates a result of least-squares fit to the form $N^*(c) = b(c^* - c)^a$, obtaining $c^* = 5.222 \pm 0.057, b = 117.8 \pm 11.5$ and $a = -2.07 \pm 0.25$. The inset shows the median running time divided by the rescaling factor $(c^* - c)^a$. All data sets $(c = 4.2, 4.3, \ldots, 5.1)$ seize to have finite size effects at a common point $N(c^* - c)^{-a} = b$, shown by the vertical line.

We can see the trend of the finite size effect from figure 5.6, which shows when the ratio $T_{\text{med}}/2N$ becomes smaller than a certain value, i.e., when the finite size effect disappears. We denote this size as N^*. The fitting presented in the figure suggests that N^* diverges at $c = 5.222 \pm 0.057$, which is in good agreement with $c_{\text{RSB}} \simeq 5.239$. Therefore, we can assume that the behavior for $c = 5.2$, although apparently seems nonlinear (slope larger than 1) in figure 5.5, is actually a finite size effect long lasting until $N \sim 10^5$ as suggested by our fitting in figure 5.6. In either case, the curves for $c > c_{\text{RSB}}$ are convex whereas the curve for $c < c_{\text{RSB}}$ are not, implying that DPLR demonstrates explicitly the RS-RSB/easy-hard correspondence.

Another point to mention is that the scaling of T_{med} shifts from N to $2N$ at $c_{\text{LR}} = 4$, consistent with the LR-core emerging. When $c_{\text{LR}} < c$, there are no vertices with degree 1 or 0, and the DPLR algorithm is forced to branch at the very beginning. For this reason, the running time T is always lower bounded by $2N$. Therefore, the asymptotic scaling of $T_{\text{med}} = 2N$ implies that the LR-core in this region is *as fragile as is could possibly be.* The scaling we saw in the previous section at $c = 2.7$ for Erdös-Rényi graphs, seems to be the same as this.

5.4.3 Conclusion and Discussions

As a conclusion, we introduced DPLR, a novel algorithm, which puts together the LR algorithm and DP. The DPLR algorithm exhibits the RS-RSB/easy-hard correspondence by exponentially improving the typical running time in the region $c_{\text{LR}} < c < c_{\text{RSB}}$ compared to the simple LR algorithm. While this further strengthens the possibility of the RS-RSB/easy-hard correspondence generally holding, our result has several other important implications as well.

First, we should emphasize that neither of the algorithms which were put together in DPLR (dynamic programming and leaf removal), were able to confirm the easy/hard threshold induced by the RS/RSB transition. Naive DP (known as branch-and-cut) already requires exponential time within the RS phase [139] for Erdös-Rényi random graphs, and naive LR fails completely in the RS phase for scale free networks. The fact that the correspondence only becomes clear after putting together these algorithms suggest the necessity of a suitable algorithm powerful enough.

Second, for most of the parameter region in the RS phase, the asymptotic scaling of T_{med} seems to be $2N$, suggesting that "easiest possible LR-cores" occur quite naturally. Furthermore, if the CBA model with $c = 5.2$ indeed has an asymptotically linear scaling of T_{med}, this means that the easy/hard transition is actually a linear/exponential time transition, more severe than polynomial/exponential. It seems that "moderately hard" LR-cores which requires nonlinear polynomial time are actually rare in naturally defined random graphs. This situation pretty much resembles that of 3-SAT, where the state-of-the-art algorithm always seems to work in linear time up until the SAT-UNSAT transition point [123]. It would be interesting to seek for natural ensembles which the degree of polynomial grows.

Thirdly, it should be emphasized that the agreement of the point where DPLR starts to take typically exponential time and the RSB transition point is nontrivial. It has been observed that "stochastic local search" algorithms frequently bypass 1RSB transitions, and are able to compute other random NP-complete problems in linear time within the 1RSB phase [140, 141]. There are also arguments on where full-RSB may be more "transparent" for algorithms [142], and the RS-RSB/easy-hard correspondence itself may be regarded nontrivial. The replica analysis obtaining the RSB transition does not rely at all on concepts such as the leaves. The DPLR algorithm on the other hand, takes advantage of the property of the MIS problem that the leaves are actually structures where the problem could be simplified. The fact that these two very different ways of analyzing the problem agree with each other on the phase transition point suggests the existence of RS-RSB/easy-hard correspondence.

Finally, our result could be seen as one example of how the NP hardness assumption we discussed in chapter 2 is physically realized, under some other assumption. If we believe that the MIS problem is indeed hard on average for some parameter regions in the random graph ensembles, this means any algorithm should be impeded in that region for some reason. Physically coupling a system with an MIS Hamiltonian to an ideal thermal bath would be one (naive) algorithm, which should be impeded as well. Our result shows that the RSB phase is responsible for the randomized MIS problem being hard on average. If we can further show this correspondence in general, the NP hardness assumption may be able to provide strategy for showing that some systems have full-RSB phases.

Furthermore, our result in this chapter becomes a guideline when we try to seek the physical consequences of the NP hardness assumption for quantum systems in the next chapter. Now that we are confident with using the DPLR algorithm as a probe for measuring average-case complexity, we can make sure that an ensemble is indeed hard on average. This point will be adopted in evaluating typical hardness of the unique solution ensemble in the next chapter.

Chapter 6

Stoquastic Quantum Annealing Applied to the Maximum Independent Set Problem

In this chapter, we consider the situation where we apply quantum annealing (QA) for the Maximum independent set (MIS) problem. As we saw in chapter 5, classical algorithms such as simulated annealing and even dynamic programming algorithms suffer from the same physical phenomena, the spin glass transition. The question of whether QA faces difficulty for the same type of physical phenomena is interesting in several ways.

First, we know that at least for some types of problems, QA can pass through energetic barriers via quantum fluctuations [143]. Thus, the spin glass picture may be completely invalid in some cases, since the picture accompanies energetic barriers between clusters in the RSB phase. To see how far this picture holds should be a general interest for statistical physicists.

Secondly, this topic would be a case where the NP-hardness assumption discussed in chapter 2 is pushed further for the quantum case. This is because we believe $\mathsf{NP} \not\subseteq \mathsf{BQP}$, and the strong (quantum) Church-Turing thesis, which implies that there *should* be some kind of an obstacle even for a quantum annealer. A similar argument in chapter 5 was mentioned in the classical case, and whether physical predictions from the NP-hardness assumption has a further physical consequence is interesting in its own right.

Finally, and relating to the first two points, is the stoquastic nature of QA. As we have seen in chapter 4, it is very likely that QA in its most basic form is not *fully* quantum owing to its stoquastic nature. So far, we do not have many *physical* properties intrinsically coming from stoquasticity, and the class of stoquastic Hamiltonians only have a rather computation-based characterization. While we may hope for a fully quantum QA to have an intrinsically quantum obstacle as a consequence of NP hardness assumption, we could seek for physical phenomena coming from stoquasticity by considering this topic.

6.1 Previous Researches

As we have introduced in chapter 3, Quantum Annealing (QA) [51, 52, 53, 75, 145] is one of the most studied strategies for utilizing quantum effects for computation. In this chapter, we will consider the stoquastic Hamiltonian in the form of Eq.(3.7)

$$\hat{H}(\lambda) = (1-\lambda)\hat{H}_{\mathrm{P}} + \lambda\hat{H}_{\mathrm{T}}, \tag{6.1}$$

where the two terms $\hat{H}_{\mathrm{P}}$ and $\hat{H}_{\mathrm{T}}$ are the problem Hamiltonian and the transverse field Hamiltonian respectively, and the parameter λ is the time-dependent control parameter.

Since the sufficient condition of how slow the quantum fluctuation should be reduced is closely related to the minimum excitation gap $\Delta E_{\min}$ of $\hat{H}(\lambda)$, QA has opened up possibility to figure out algorithmic difficulties from physical perspectives. As we discussed in section 3.1, the quantum adiabatic theorem [144] guarantees that the quantum state starting from the ground state of $\hat{H}_{\mathrm{T}}$ will always remain in the instantaneous ground state with high probability, as long as λ is changed using time longer than $O(\Delta E_{\min}^{-2})$. Thus, for a fixed problem instance embedded in $\hat{H}_{\mathrm{P}}$, if there are only polynomially small energy gaps, it implies that QA can solve the specific problem in polynomial time, which is considered as "efficient" in the computer science community [18, 17].

Surprisingly, early studies hinted polynomial time computation of NP-hard problems [146], however they turned out to be artefacts of finite-size effects [147]*1. As we believe that NP$\not\subseteq$BQP from oracle results[49], or equivalently from the NP hardness assumption [34], we expect for quantum annealers to fail for NP-hard problems, indifferent to the fact that it is only BstoqP and may not be BQP-complete. This could be seen as a *computational complexity based* physical conjecture that says if a Hamiltonian set corresponds to an NP-hard problem, then Hamiltonians with exponentially small energy gaps should exist, especially when the solution is unique. Providing a physical picture to this conjecture would be of great importance for the understanding of the connection between computational complexity and physics. One of the convincing arguments was made using numerical approaches [71], showing that a fraction of samples exhibit first order phase transition-like behaviors in terms of the usual spin-glass order parameter q. Since first order transitions in quantum spin systems are usually accompanied with an exponentially small energy gap [148], they provide good reason that NP-hard problems could not be computed in polynomial time using QA. However, those first order transitions are strongly sample dependent, and no concrete arguments have been made for explaining the phenomena with a physical mechanism.

6.1.1 The Spin-glass picture

For the quantum random energy model (QREM) [149], it is known that there is a 1RSB phase transition, and that transition accompanies an exponentially small energy gap. However, QREM is an extremely simplified model, with somewhat pathologic properties. It is also a "trivially exponentially hard problem" when considered as an optimization problem, and does not suffice for our motivation of seeking the physical consequences of the NP hardness assumption.

For some NP-hard models, it is known that a spin-glass transition occurs. There are two different ways of exponentially small energy gaps emerging. One is where the exponentially small energy gap occurs at the spin-glass phase transition point, such as in ref. [101]. Another way is that the spin-glass phase transition point itself only has a polynomially small energy gap, and *within* the spin-glass phase, there exists an exponentially small energy gap. The latter pattern was rigorously shown to be true for the Hopfield model [150], which is again a relatively simple model compared to NP-hard problems. Also, the "spin-glass transition" which the quantum Hopfield model here experiences is always in the RS phase, which is different from that of NP-hard problems.

Therefore, it remains open if NP-hard problems in general undergo the same physical scenario of a spin-glass transition.

*1 This claim would have been immediately disgarded if the NP hardness assumption was considered.

6.1.2 The Localization Picture

It was argued in [152], that the Anderson/many-body localization transition happens during the QA process. While the energy gap for the localization transition point is not discussed, it is argued that *within* the localized phase, avoided crossing with exponentially small energy gaps occur [151]. Their argument essentially relates the stoquastic annealing Hamiltonian Eq.(6.1) and the tight-binding model for Anderson localization [153] (AL) as

$$(1-\lambda)\hat{H}_{\mathrm{P}} - \lambda \sum_i \hat{\sigma}_i^x \quad = \quad (1-\lambda) \sum_{\boldsymbol{\sigma}} H_{\mathrm{P}}(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle\langle\boldsymbol{\sigma}| - \lambda \sum_{\boldsymbol{\sigma},\boldsymbol{\sigma}':\mathrm{n.n.}} |\boldsymbol{\sigma}'\rangle\langle\boldsymbol{\sigma}|, \tag{6.2}$$

where the sum in the second term takes all nearest neighbor pairs of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$. Apparently, the expression on the right hand side resembles the Hamiltonian of AL, but there are caveats.

First, while the second term corresponds to the hopping term of AL, it differs in the sense that the hopping occurs in configuration space and not real space. Since the existence of AL phase crucially relies on the space dimension, we are not sure if AL occurs in our setting as well. This is because the configuration space we are considering now is an N-dimensional hypercube with $N \to \infty$ in the thermodynamic limit. Secondly, while the AL model has random on-site potential, usually drawn from i.i.d. uniform distribution, the on-site potential of the above model is simply the cost function, which is obviously a *random instance*, but has correlations and is far from independent.

For these reasons, it is not clear how much the analogy with AL discussed in [152] is valid. No concrete demonstration of the AL phase in QA of NP-hard problems has been found so far.

6.1.3 Our Strategy

Both of the above scenarios explain the first order transitions as a phenomenon within a specific underlying phase. However, while there are arguments against the localization picture [154], studies against the quantum spin-glass picture [155] exist as well, and the understanding for the phase transition inducing the first order transitions in [71] is very much unsettled.

In this work, we see if the "underlying phase transition" scenarios are correct at all, for the unique solution ensemble using quantum Monte Carlo (QMC) simulations. A natural strategy to see the underlying phase transition would be to take the sample average (disorder average) of the spin-glass order parameter q which exhibited a first order transition-like behavior in individual samples. However, we find that this quantity does not exhibit any singularities when the sample average $\langle \bar{q} \rangle$ is taken. The implication is not only that there is no spin-glass transition, but also that another measure is necessary to see the phase transition in question, if it exists. We thus used the notion of *fidelity susceptibility* [156, 157, 158]. This quantity quantifies how rapid the ground state is changing in the λ direction. The fidelity susceptibility χ_{F} is also proportional to the symmetric logarithmic derivative (SLD) Fisher information metric [159], and has been recently under interest for detecting phase transitions where the order parameter is unknown, such as topological order phases. Our work suggests that χ_{F} is also useful for quantum spin-glass-like models with frustrated quenched disorders.

By using stochastic series expansion (SSE), a variant of quantum Monte Carlo methods, we estimate the sample average of the fidelity susceptibility $\overline{\chi_{\mathrm{F}}}$. We show for a specific NP-hard problem that QA undergoes a phase transition at a certain value of λ. At this transition point, $\overline{\chi_{\mathrm{F}}}$ diverges, while other common quantities such as $\bar{q}$ do not show or has very weak singularity. This implies that although there is a quantum phase transition at the value of λ, the order parameter for this transition is yet unknown. However, since all of the first order like transitions occur at values below the transition point, this result suggests that the first order transitions could be understood as a phenomenon within a non-trivial quantum phase.

6.2 The Setup

6.2.1 The Model

We fix the NP-hard problem in consideration to the Maximum Independent Set (MIS) problem. As we have introduced in chapter 5, MIS is the problem where given a graph $G = (V, E)$, one finds the maximum subset of vertices $I^* \subset V$ such that no two vertices in I^* are adjacent (i.e. $\forall i, j \in I^*, (i,j) \notin E$). Finding the solution to the MIS problem is equivalent to finding the ground state of a Hamiltonian with the Pauli matrix in the form of

$$\hat{H}_{\mathrm{P}} = \frac{\alpha}{4} \sum_{(v,w)\in E} \hat{\sigma}_v^z \hat{\sigma}_w^z - \sum_v \frac{2 - \alpha d_v}{4} \hat{\sigma}_v^z, \tag{6.3}$$

which is the same as Eq.(), but in the quantum form. We will generate random instances of the MIS problem by randomly generating the graph G in target. We discuss how the random ensembles are ensured to be hard on average in the next section.

6.2.2 The Ensemble

In this chapter, we consider two different ensembles. One is the unique solution ensemble and the other is the Erdöes-Rényi ensemble. The focus is more on former, since it has a direct prediction from the NP hardness assumption as we will see.

The Erdöes-Rényi Ensemble

The Erdöes-Rényi ensemble (ERE) is simply the Erdöes-Rényi random graphs [160] which we also considered in chapter 5. We have already seen that ERE enters a full-RSB phase when the average degree c is larger than $e = 2.718\ldots$ [130], and this becomes an important point when we motivate studying ERE later. Furthermore, from our results in chapter 5, we know that the MIS problem on ERE becomes hard on average in the RSB phase. This also is important when we construct the other ensemble with a unique solution.

An important difference with the unique solution ensemble is that an Erdöes-Rényi random graph typically has multiple (exponentially many) MIS, known as residual entropy. As we will explain in detail in the next subsection, this disables defining the computation time for QA in the most straightforward theoretical way. Thus in our study, ERE plays a role of supporting our claims from the statistical physics side, rather than being the main subject.

The Unique Solution Ensemble

The unique solution ensemble (USE)*2 is our main focus in this chapter.

If there were multiple solutions, two problems arise. One is that the minimum energy gap $\Delta E_{\min}$ becomes ill-defined, since the $\hat{H}_{\mathrm{P}}$ has a degenerate ground state causing

*2 This ensemble is often referred as the "unique solution assignment", and is abbreviated as USA in literature, e.g., [71].

$\Delta E_{\min} \to 0$ at $\lambda \to 0$. The other problem is more subtle. If there were multiple solutions, an exponentially small energy gap would no longer imply exponential computation time for QA. This is because even if we have a Landau-Zener transition at the point of exponentially vanishing gap, if the state remains in the first excited state $|\Psi_1(\lambda)\rangle$ for the rest of the annealing schedule, we will end up in the ground state anyway. Therefore, having a unique solution is essential for having a well-defined computation time.

Another reason we consider USE is that the first order transitions mentioned in section 6.1 are previously only observed in those examples which have unique solutions[71].

We construct USE by randomly adding extra edges to an Erdös-Rényi random graph. The details of this procedure are explained in Appendix A. The important point is that the procedure is probabilistic, which gives us a distribution over MIS instances with a specific size.

We should make a final remark on the average case hardness of USE. It is crucial to ensure that the process of making the solution unique does not make the complexity of the problem easy. The intuition is that adding edges to make the solution unique only effectively increases the average degree, and pushes the graph deeper into the RSB/hard phase. We describe the details of how this point is made sure in Appendix B.

6.2.3 The Method: SSE and EMC

We adopt the SSE method for Ising spin systems [95, 162] as explained in details in section 4.1.3. The SSE method effectively takes the Trotter limit in the path integral Monte Carlo method [86, 161], and is therefore free from systematic error caused by the Trotter decomposition. We decompose the Hamiltonian of Eq.(6.1) in the same manner as in section 4.1.3, and use the z basis $\{|\boldsymbol{\sigma}\rangle\}$. Here we decouple the Hamiltonian as $\hat{H}(\lambda) = -\sum_k \hat{W}_k$. We add constant terms appropriately as explained in section 4.1.3, so that for all $|\boldsymbol{\sigma}\rangle$ and $\hat{W}_k$, $\hat{W}_k|\boldsymbol{\sigma}\rangle = W_{k,\boldsymbol{\sigma}}|\boldsymbol{\sigma'}\rangle$ with $W_{k,\boldsymbol{\sigma}} \geq 0$. Here, we have the interaction terms $(1-\lambda)J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z$and local field terms $(1-\lambda)h_i\hat{\sigma}_i^z$ coming from $\hat{H}_\mathrm{P}$, and transverse field terms $\lambda\hat{\sigma}_i^x$ coming from $\hat{H}_\mathrm{T}$. The partition function Z is expanded as

$$Z := \mathrm{Tr}[e^{-\beta\hat{H}}] \tag{6.4}$$

$$= \sum_{n=0}^{\infty} \frac{\beta^n}{n!} \sum_{\{k_l\}} \sum_{\boldsymbol{\sigma}} \langle\boldsymbol{\sigma}| \prod_{l=1}^{n} \hat{W}_{k_l} |\boldsymbol{\sigma}\rangle, \tag{6.5}$$

where β is the inverse temperature. SSE samples terms in the above summation using Markov chain Monte Carlo methods, by changing $|\boldsymbol{\sigma}\rangle$, n and $\{\hat{W}_{k_l}\}_l$.

Furthermore, we also adopt the exchange Monte Carlo method (parallel tempering) to accelerate equilibration [99, 102] as discussed in section 4.2.2. We divide the parameter region $[\lambda_\mathrm{low}, \lambda_\mathrm{high}]$ into $R-1$ equidistributed intervals, and run R different SSE simulations with the corresponding λ_r. By applying Eq.(4.32), we get the exchange probability for SSE, of adjacent λ_r,

$$P_{\lambda_r \leftrightarrow \lambda_{r+1}} = \min\left[1,\ \left(\frac{\lambda_r}{\lambda_{r+1}}\right)^{k_{r+1}-k_r} \left(\frac{1-\lambda_r}{1-\lambda_{r+1}}\right)^{l_{r+1}-l_r}\right],$$

where k_r and l_r denote the number of operators coming from $\hat{H}_\mathrm{T}$ and $\hat{H}_\mathrm{P}$, respectively.

As discussed in section 4.2.1, we take the strategy of taking the inverse temperature β large enough so that $\beta > \Delta E_1 = E_1 - E_0$, letting the equilibrium state effectively the same as the ground state. Practically, we sample the equilibrium state of $\beta = 3.5N$ with N being the number of vertices of the problem. By fixing the system size N and increasing

β, we see that usual observables saturate around $\beta \sim 1.5N$ and $\overline{\chi_{\rm F}}$ also saturates at $\beta \sim 3.5N$. Thus, sampling the thermal equilibrium state at $\beta = 3.5N$ is sufficient to see the properties of the ground state except for the transition points. The arguments for the transition points will be discussed later. 3072 samples (instances) were taken for system sizes $N = 20$ and 30, and 1.6×10^5 to 1.28×10^6 MCS to measure the quantities after the same amount for equilibration. In the following, we show 256 samples with $N = 50$ since they exhibit clearer first order transitions, but will not use them for sample averages due to lack of enough samples.

6.2.4 Measured Observables

Before we explain our results, we will first introduce the physical quantities we measured in our study. We also explain how to measure the quantities using SSE.

The Overlap Parameter q, and its Derivative

The overlap parameter q (also called as the spin-glass order parameter) is defined as [71]

$$q := \frac{1}{N}\sum_i \langle \hat{\sigma}_i^{(1)} \hat{\sigma}_i^{(2)} \rangle = \frac{1}{N}\sum_i \langle \hat{\sigma}_i \rangle^2, \tag{6.6}$$

where the upper suffixes (1) and (2) are the labels for two independent systems with the same quenched disorder. Note that this definition is the *expectation value* of the overlap parameter q we defined in Eq.(5.5), thus a notation as $\langle q \rangle$ may be more suitable. Caution is needed since previous studies on QA with glassy systems [101, 71] focus on this quantity, and denote it as q instead of $\langle q \rangle$. In this chapter, we may also write q for $\langle q \rangle$ according to convention. Although q does not contain the entire information of $P(q)$, it captures singularities in $P(q)$, for instance an RSB transition. If the structure of $P(q)$ becomes non-trivial at some point, with more than one delta peak, the average value $\langle q \rangle = \int qP(q)dq$ will also have a singularity, at least in the derivative by λ. Thus, measuring q is effective for detecting the spin-glass phase transition, and is simple since we only have to measure the average magnetization $\langle \hat{\sigma}_i^z \rangle$ for each site i.

The above argument encourages us to also measure the derivative $\partial q/\partial \lambda$. Measuring the first derivative of q

$$\frac{\partial q}{\partial \lambda} = \frac{1}{N}\sum_i 2\langle \hat{\sigma}_i^z \rangle \frac{\partial \langle \hat{\sigma}_i^z \rangle}{\partial \lambda}, \tag{6.7}$$

reduces to measuring

$$\frac{\partial \langle \hat{\sigma}_i^z \rangle}{\partial \lambda} = \frac{1}{Z}\frac{\partial \mathrm{Tr}[e^{-\beta \hat{H}(\lambda)} \hat{\sigma}_i^z]}{\partial \lambda} - \frac{1}{Z}\frac{\partial Z}{\partial \lambda}\langle \hat{\sigma}_i^z \rangle. \tag{6.8}$$

Expanding Z the same way for deriving SSE as in Eq.(4.15), we obtain

$$\frac{1}{Z}\frac{\partial Z}{\partial \lambda} = \frac{1}{Z}\frac{\partial}{\partial \lambda}\sum_{\boldsymbol{\sigma}}\sum_{n=0}^{\infty}\frac{\beta^n}{n!}\sum_{k(l)}\langle \boldsymbol{\sigma}|\,(1-\lambda)^l\,\lambda^k \prod_{l=1}^{n}\hat{W}'_{k(l)}|\boldsymbol{\sigma}\rangle \tag{6.9}$$

$$= \frac{1}{Z}\sum_{\boldsymbol{\sigma}}\sum_{n=0}^{\infty}\frac{\beta^n}{n!}\sum_{k(l)}\langle \boldsymbol{\sigma}|\left(\frac{-l}{1-\lambda}(1-\lambda)^l\lambda^k + \frac{k}{\lambda}(1-\lambda)^l\lambda^k\right)\prod_{l=1}^{n}\hat{W}'_{k(l)}|\boldsymbol{\sigma}\rangle \tag{6.10}$$

$$= \left\langle \frac{k}{\lambda} - \frac{l}{1-\lambda} \right\rangle_{\rm config}, \tag{6.11}$$

where we borrow the notation $\langle\cdot\rangle_{\text{config}}$ from section 4.1.3. Here, l and k denote the number of operators coming from $\hat{H}_{\mathrm{P}}$ and $\hat{H}_{\mathrm{T}}$ respectively, and $\hat{W}'$ are the operators without the coefficients λ or $(1-\lambda)$. Similarly it could be calculated that,

$$\frac{1}{Z}\frac{\partial \mathrm{Tr}[e^{-\beta\hat{H}(\lambda)}\hat{\sigma}_i^z]}{\partial\lambda} = \left\langle \left(\frac{k}{\lambda} - \frac{l}{1-\lambda}\right)\hat{\sigma}_i^z \right\rangle_{\text{config}}. \tag{6.12}$$

Therefore,

$$\frac{\partial q}{\partial \lambda} = \frac{2}{N}\sum_i \langle\hat{\sigma}_i^z\rangle \left\{ \left\langle \left(\frac{k}{\lambda} - \frac{l}{1-\lambda}\right)\hat{\sigma}_i^z \right\rangle - \langle\hat{\sigma}_i^z\rangle \left\langle \frac{k}{\lambda} - \frac{l}{1-\lambda} \right\rangle \right\}. \tag{6.13}$$

Similarly, derivative of other quantities tend to become related to covariances of simple quantities.

The Answer Fidelity F_{ans}

The notion of fidelity was introduced in order to address the "closeness" of two quantum states. Naively thinking, we can simply use the inner product of two pure states, and in fact the absolute value of it turns out to be the fidelity for pure states (or the square of it depending on style). Here we define the fidelity between two quantum states $|\psi\rangle$ and $|\phi\rangle$ as

$$F(\psi,\phi) = \big|\langle\psi|\phi\rangle\big|. \tag{6.14}$$

Since we take the strategy of measuring the finite temperature state of large enough β to obtain the ground state, a generalization of the above simple notion of fidelity to mixed states is necessary. The most straight forward way to understand fidelity would be to set some axioms which a closeness function should satisfy. We choose the axioms as below.

1. $0 \le F(\psi,\phi) \le 1$ (normalization)
2. $F(\psi,\phi) = F(\phi,\psi)$ (symmetric)
3. $F(\hat{U}\psi,\hat{U}\phi) = F(\psi,\phi)$ (unitary invariant)
4. $F(\psi_1\otimes\psi_2,\phi_1\otimes\phi_2) = F(\psi_1,\phi_1)F(\psi_2,\phi_2)$ (multiplicity)

It is known that the definition of fidelity for mixed quantum states

$$F(\hat{\rho},\hat{\rho}') = \mathrm{Tr}\sqrt{\hat{\rho}^{1/2}\hat{\rho}'\hat{\rho}^{1/2}}, \tag{6.15}$$

satisfies the above and is the unique function up to arbitrary powers. Note that if we assume pure states for the density operators, we recover Eq.(6.14). Here, we define the root of an operator $\sqrt{\hat{\rho}} = \hat{\rho}^{1/2}$ as the unique operator $\hat{X}$ which is positive-semidefinite and satisfies $\hat{X}^2 = \hat{\rho}$. This is unique when $\hat{\rho}$ itself is positive semidefinite, which is always the case when $\hat{\rho}$ is actually an density operator. The fact that the above definition satisfies all of the axioms should be not hard to confirm. It can be rewritten in the form

$$F\left(\frac{e^{-\beta\hat{H}(\lambda)}}{Z(\lambda)}, \frac{e^{-\beta\hat{H}(\lambda')}}{Z(\lambda')}\right) = \frac{\mathrm{Tr}\sqrt{e^{-\beta\hat{H}(\lambda)/2}e^{-\beta\hat{H}(\lambda')}e^{-\beta\hat{H}(\lambda)/2}}}{\sqrt{Z(\lambda)Z(\lambda')}} \tag{6.16}$$

$$= \frac{\mathrm{Tr}[e^{-\beta\hat{H}(\lambda)/2}e^{-\beta\hat{H}(\lambda')/2}]}{\sqrt{Z(\lambda)Z(\lambda')}} \tag{6.17}$$

if the mixed states are canonical states and $-\beta\hat{H}$ is positive semidefinite. The positive semi-definiteness is always satisfied if we are considering stoquastic Hamiltonians*3. Accordingly, to obtain Eq.(6.17), we use the relation

$$\mathrm{Tr}[\hat{X}^\dagger \hat{X}] = \mathrm{Tr}[\hat{X}^2] \tag{6.18}$$

which holds when $\hat{X}$ is real and symmetric.

We measure the fidelity of the ground state $|\Psi_0(\lambda)\rangle$ with the answer state $|\boldsymbol{\sigma}_{\mathrm{ans}}\rangle = |\Psi_0(\lambda = 0)\rangle$. This is done easily, by simply measuring $\langle \hat{P}_{\mathrm{ans}}\rangle$ where $\hat{P}_{\mathrm{ans}} = |\boldsymbol{\sigma}_{\mathrm{ans}}\rangle\langle\boldsymbol{\sigma}_{\mathrm{ans}}|$.

The Fidelity Susceptibility χ_{F}

Both of the spin-glass order parameter q and the answer fidelity F_{ans}, turns out to be a quantity which does not capture the *underlying* phase transition, as we will see in the following section. The *fidelity susceptibility* is a notion which enables us to capture phase transitions when we do not know the order parameter. The idea is to see the fidelity between two ground states with close λ. If the two states are both in the same phase, the fidelity susceptibility should be close to 1, while if they are separated by a phase transition in between, the fidelity should drop. This leads us to take the derivative of the fidelity.

However, there are subtleties to be taken care of. First of all, since the fidelity is defined as a function with *two* quantum states as inputs, we have two parameters which could be varied. Let us consider the case where we fix one state and vary the other, as $|\langle\Psi_0(\lambda)|\Psi_0(\lambda+\epsilon)\rangle|$. Now, if we take the first derivative by ϵ

$$\left|\frac{\partial}{\partial\epsilon}\langle\Psi_0(\lambda)|\Psi_0(\lambda+\epsilon)\rangle\right| = \left|\langle\Psi_0(\lambda)|\frac{\partial|\Psi_0(\lambda+\epsilon)\rangle}{\partial\lambda}\right|, \tag{6.19}$$

the result is always 0. This is because from

$$0 = \frac{\partial}{\partial\lambda}\langle\Psi_0(\lambda)|\Psi_0(\lambda)\rangle = \langle\Psi_0(\lambda)|\frac{\partial|\Psi_0(\lambda)\rangle}{\partial\lambda} + \frac{\partial\langle\Psi_0(\lambda)|}{\partial\lambda}|\Psi_0(\lambda)\rangle, \tag{6.20}$$

the term is always purely imaginary. Since we are now considering a ground state of a stoquastic Hamiltonian this term should be real as well, so the only possibility is 0 *4. This corresponds to the graph of $|\Psi_0(\lambda)|\Psi_0(\lambda+\epsilon)|$ having a smooth curve as a function of ϵ and thus having no slope at $\epsilon = 0$.

Therefore, we take the *second* derivative of the fidelity which quantifies how sharp the peak of the fidelity as a function of ϵ is. This corresponds to how rapidly the ground state is changing in the Hilbert space. Thus, we define the fidelity susceptibility as

$$\chi_{\mathrm{F}}(\lambda) := -\left.\frac{\partial^2 \log F(\Psi_0(\lambda), \Psi_0(\lambda+\epsilon))}{\partial\epsilon^2}\right|_{\epsilon=0}. \tag{6.21}$$

Measuring the fidelity susceptibility using QMC is not straight forward. In order to measure the fidelity susceptibility, we use the finite temperature fidelity as in [163],

$$F(\lambda, \lambda+\epsilon) = \sqrt{\frac{\mathrm{Tr}[e^{-\beta\hat{H}(\lambda)/2}e^{-\beta\hat{H}(\lambda+\epsilon)/2}]}{(\mathrm{Tr}[e^{-\beta\hat{H}(\lambda)}]\mathrm{Tr}[e^{-\beta\hat{H}(\lambda+\epsilon)}])^{1/2}}}, \tag{6.22}$$

which corresponds to the *square root* of the fidelity we defined in Eq.(6.15). The difference in the power for the definition of fidelity does not change the interpretation of the

*3 The definition of $\hat{O}$ begin semi positive-definite is to have $\langle\psi|\hat{O}|\psi\rangle \geq 0$ for arbitrary $|\psi\rangle$. This is exactly the condition for SSE to be able to applied.

*4 Even for non-stoquastic systems, the imaginary term can be erased by a global phase shift $e^{-i\theta}$.

fidelity susceptibility as the quantification of how rapid the ground state is changing. By expanding Eq. (6.22) up to the second term in ϵ, we obtain the representation of χ_{F} for finite temperature as

$$\chi_{\mathrm{F}}^{T\neq 0} = \frac{\langle k_L k_R\rangle - \langle k_L\rangle\langle k_R\rangle}{2\lambda^2} + \frac{\langle l_L l_R\rangle - \langle l_L\rangle\langle l_R\rangle}{2(1-\lambda)^2} - \frac{\langle k_L l_R\rangle - \langle k_L\rangle\langle l_R\rangle}{2\lambda(1-\lambda)} - \frac{\langle l_L k_R\rangle - \langle l_L\rangle\langle k_R\rangle}{2\lambda(1-\lambda)}, \tag{6.23}$$

where k_* and l_* denote the number of operators within the left or right half (depending on the subindex) of the operator string $\{\hat{W}_{k_l}\}_{l=1}^{n}$, coming from $\hat{H}_{\mathrm{T}}$ and $\hat{H}_{\mathrm{P}}$, respectively. $\langle\cdots\rangle$ represents the Monte Carlo average. The center of the string is determined probabilistically for every sampled configuration, according to the binomial distribution among the possible $n+1$ points of division.

6.3 Results on the Unique Solution Ensemble

Equipped with the above quantities and the methods to measure them, we see if there are physical phase transitions in the MIS problem with unique solutions. Note that as we discussed in section 6.2.2 and in Appendix B, the USE is a random distribution over the MIS problem which is *typically hard.* Our aim is to see physical phenomena which could be seen as the "physical consequence of the NP hardness assumption", and further confirm scenarios regarding "underlying phase transitions" which were previously proposed.

6.3.1 Spin-glass Order Parameter q

The spin-glass order parameter q (the overlap parameter) was used to capture the "first order phase transitions" in [71]. Our result presents a similar behavior and we explain this in detail. What we see immediately is that q behaves very differently for different samples (Fig. 6.1). Not all, but some samples exhibited an acute increase in q after a characteristic dip with decreasing λ, which we show in red lines in the figure.

As we discussed in section 5.2.1, the overlap parameter q quantifies how "frozen" the spins are. We can immediately see from the definition Eq.(6.6) that $q(\lambda)$ is 0 at $\lambda = 1$ and is 1 at $\lambda = 0$ since the solution is unique. During the process of decreasing the transverse field, the spins are expected to freeze due to the reducing of quantum fluctuations. Therefore, the "dip" of q, which is a non-monotonic behavior seems counterintuitive. This could be understood as follows. Let's assume that the ground state experiences a first order phase transition. By this, we mean that the dominant configuration before and after the transition, $|\boldsymbol{\sigma}_{\mathrm{before}}\rangle$ and $|\boldsymbol{\sigma}_{\mathrm{after}}\rangle$, has a Hamming distance of $O(N)$ In this case, there are $O(N)$ qubits which differ between $|\boldsymbol{\sigma}_{\mathrm{before}}\rangle$ and $|\boldsymbol{\sigma}_{\mathrm{after}}\rangle$, which cancels each other for contributing to q in the form $\langle\hat{\sigma}_i^z\rangle^2$ at the phase transition point. This would result in a decrease of $O(1)$ for q, which is a dip.

The above explanation with first order transition for the dip is obviously not decisive on its own. It is unclear if the two ground states before and after the transition could be written with dominant z basis configurations[*5]. It also only provides a sufficient condition. If there really *is* a first order transition of that kind, there would indeed be a dip in q

[*5] Of course, if there *is* a localization transition for this model, and the first order transitions take place within the localized phase, the argument will be completely justified. The problem is that whether if the "underlying phase transition" scenario (including localization) really holds or not was the main purpose of our study!

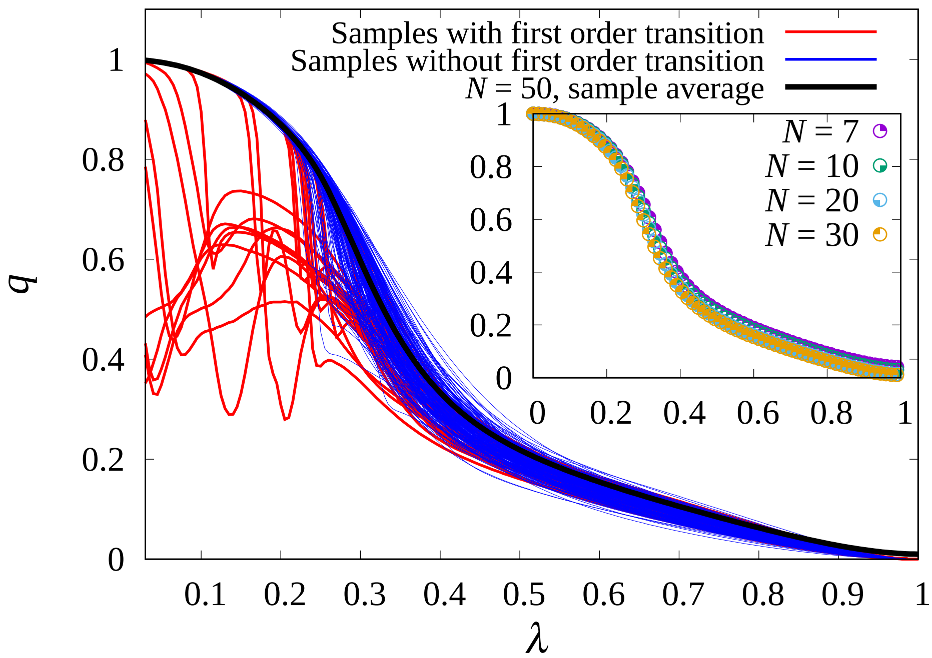

Fig. 6.1: The overlap parameter q is drawn as a function of λ for 256 samples with $N = 50$. 27 samples with dips are drawn in red, and the remaining 229 samples without dips are in blue. The black line is the sample average $\bar{q}$. The inset presents $\bar{q}$ for different sizes.

accordingly to the argument. However the existence of a dip may be explained with another physical picture, and is not necessarily an evidence for the first order transition. Furthermore, strictly speaking, first order phase transitions are notions which are well-defined only in the thermodynamic limit.

However we have several other supporting evidence for the dips indicating first order transition. One is that double peaks in the histogram of q and F_{ans} indicating phase coexistence were confirmed at the "transition points" of those "first order transitions" [165]. We will also see a finite jump for F_{ans} at the same points, further indicating a sudden change to the $|\boldsymbol{\sigma}_{\text{ans}}\rangle$ dominant ground state. Therefore, we will use the term "first order transition" for this phenomena in this thesis, assuming the first order phase transition picture holds. In this case, there would be exponentially small energy gaps at the first order transition points, from the Landau-Zenner argument [71, 152]. More precisely, in order to have two distant states $|\boldsymbol{\sigma}_{\text{before}}\rangle$ and $|\boldsymbol{\sigma}_{\text{after}}\rangle$, to have avoided crossings[*6], the $O(N)$ th perturbation is needed, resulting in the exponentially small energy gap.

Now that we have evidence of exponentially small energy gaps, and a physical picture of first order transitions, can we go further and understand the first order transitions *as a whole*? As we can see from Fig. 6.1, the transition points and even the presence of the first order transitions are sample dependent. This means that we cannot argue simply that there is always a first order transition accompanying an exponentially small energy gap at a specific value of λ.

The sample dependency of q itself is rather consistent with the average case complexity

[*6] Remember that degeneracy is generally prohibited due to the non-crossing theorem (thm. 4) explained in section 3.2.

of the MIS problem with this particular ensemble, since not all instances need to be hard. We can say that the samples with first order transitions are the hard instances while the samples without the first order transitions are the easy instances. Since our model is set to be hard on average, we expect that the proportion of the instances with first order transitions should grow in the thermodynamic limit. This tendency was indeed observed for another NP-complete model, however we could not conclude so for our model as well, due to limitation of size. We only see a slight increase in the ratio of instances with first order transition. This is partly because we only start to observe first order transitions at $N \gtrsim 25$. This fact implies that it is necessary to consider size somewhat larger than the size accessible by the state-of-the-art diagonalization methods, in order to access the essential hardness QA faces when tackling computationally hard problems.

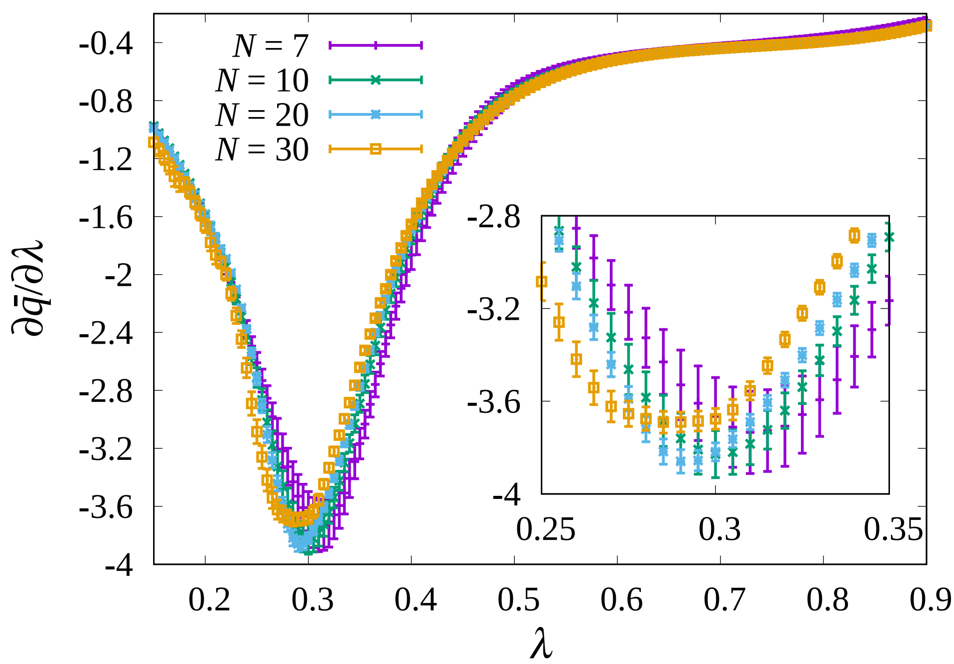

Fig. 6.2: The λ derivative of the averaged spin-glass order parameter $\bar{q}$ for different sizes are drawn for different sizes. The inset is a magnified view at the minimum point.

We can also see that the first order transitions always take place in the region $\lambda \lesssim 0.3$. This hints that there may be an underlying phase transition around $\lambda \sim 0.3$, which the low-λ phase induces the first order transitions. Furthermore, it is known that within the RSB phase of spin glasses, self-averaging is broken, meaning that values of certain quantities such as $P(q)$ do not match for all instances even in the thermodynamic limit [166][*7] ($P(q)$ as we explained in section 5.2.1 is defined by taking average over all samples.). This leads us to an expectation that the severe sample dependence of q at the "low λ region" is an appearance of the breaking of self-averaging.

However, when we take the sample average of q to see the behavior of the ensemble as in Fig. 6.1 (inset), we are unable to see any size dependence nor singularities. The derivative with respect to λ, $\partial q/\partial\lambda$, also seems to have no singularities as seen in Fig. 6.2. This

[*7] Thermodynamic quantities such as energy, free energy, entropy, etc... do have self-averaging properties, and an instance will have the same value as the random average for those quantities in the thermodynamic limit.

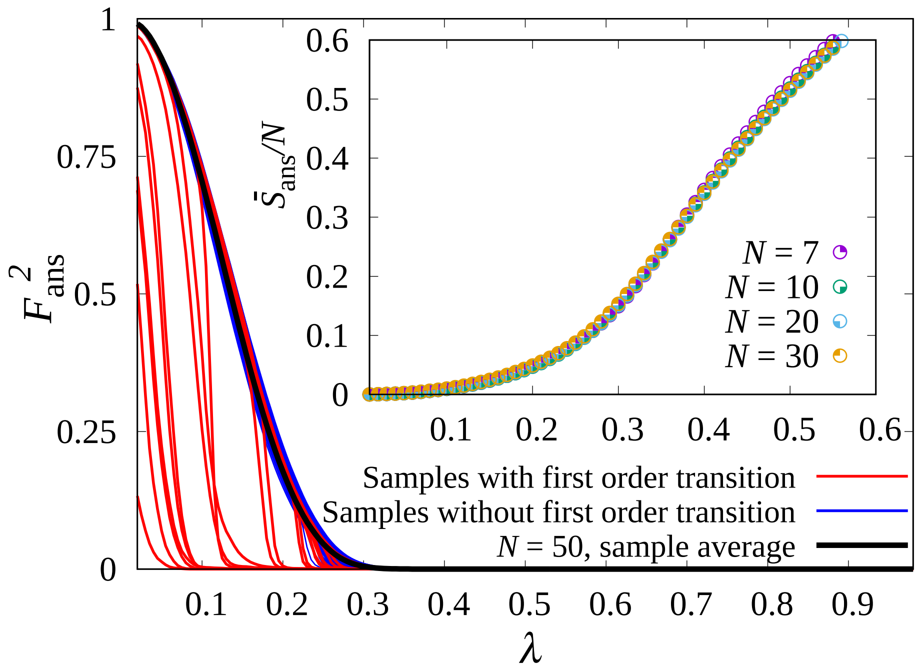

Fig. 6.3: Square of the answer fidelity $F_{\text{ans}}^2 = \langle \hat{P}_{\text{ans}} \rangle$ for 256 samples with $N = 50$. The inset shows the sample average of $-\log_2 \langle \hat{P}_{\text{ans}} \rangle / N$, i.e. $\overline{S_{\text{ans}}}/N$ for different sizes.

implies that although q was a good quantity for detecting the first order phase transitions for individual samples in QA, it does not capture the underlying phase transition for this problem. This suggests that there is no RSB phase transition as discussed in section 5.2.1, since RSB would imply singularity in q, which is not observed.

6.3.2 Answer Fidelity F_{ans}

Provided that there is no underlying transition detectable by q, let us consider checking the localization scenario.

As we explained briefly in section 6.1.2, the localization takes place in the configuration space [152], which could be described as an N-dimensional hypercube, each vertex corresponding to one configuration, i.e. a particular product state in the z-direction. The ground state is fully extended (has equal weight on all 2^N states) at $\lambda = 1$, and since our model has only one solution, at $\lambda = 0$ the state is fully localized. An easily measured quantity that quantifies this situation is the fidelity between the ground state of a particular λ and that of $\lambda = 0$. We will call this as the answer fidelity $F_{\text{ans}} := F(\lambda, 0)$, and this will represent the probability of observing the correct answer by a projection measurement with the z-basis, $\langle \hat{P}_{\text{ans}} \rangle$.

Since the answer fidelity is 0 for $\lambda = 1$ in the thermodynamic limit and is 1 for $\lambda = 0$, it could be a natural candidate for an order parameter. Indeed, as we can see from Fig. 6.3, the answer fidelity jumps from almost 0 to a finite value in samples exhibiting first order transition, at the same value of λ as the dips in q. This strongly suggests that the first order transition could be understood as a transition into a $|\boldsymbol{\sigma}_{\text{ans}}\rangle$ dominant state from a completely different state, enforcing the arguments for the exponentially small energy gap associated with it.

Furthermore, samples without first order transitions start to have non-negligible values from a certain value of λ. It is tempting to think that there is an underlying phase transition which is a transition from a $F_{\text{ans}} = 0$ phase to a $F_{\text{ans}} > 0$ phase, with the hard instances somehow gets stuck in another basin. However, this turns out to be incorrect. This is because F_{ans} seems to converge to 0 for all $\lambda > 0$. In fact, if we plot $S_{\text{ans}} := -\log_2\langle\hat{P}_{\text{ans}}\rangle$ normalized by N, they collapse into a common curve as in the inset of Fig. 6.3. This is natural, since when λ is small enough and if F_{ans} is larger than the fidelity between the ground state and any other basis state of the z-direction $|\boldsymbol{\sigma}\rangle$, S_{ans} is actually equivalent to the Rényi entropy [167, 168]

$$S_n(\lambda) := \frac{1}{1-n}\sum_{\boldsymbol{\sigma}}\left|\langle\boldsymbol{\sigma}|\mathrm{GS}(\lambda)\rangle\right|^{2n}, \tag{6.24}$$

in the $n \to \infty$ limit, where $|\mathrm{GS}(\lambda)\rangle$ is the ground state of $\hat{H}(\lambda)$. By assuming that S_∞/N converges to a finite value, and that $S_\infty = S_{\text{ans}}$, it is easy to see that F_{ans} goes to 0 for all $\lambda > 0$.

Similarly to q, although both F_{ans} or S_{ans} seems to capture the first order transitions, when we take the sample average of them, no singularities could be observed. Furthermore, when $n = 2$, the Rényi entropy becomes the logarithm of the inverse participation ratio, a frequently used quantity to see localization [169]. The absence of singularity in the F_{ans} or S_{ans} strongly suggests that the original Anderson localization scenario [152] also fails to show an underlying phase transition of our model.

6.3.3 Fidelity Susceptibility χ_{F}

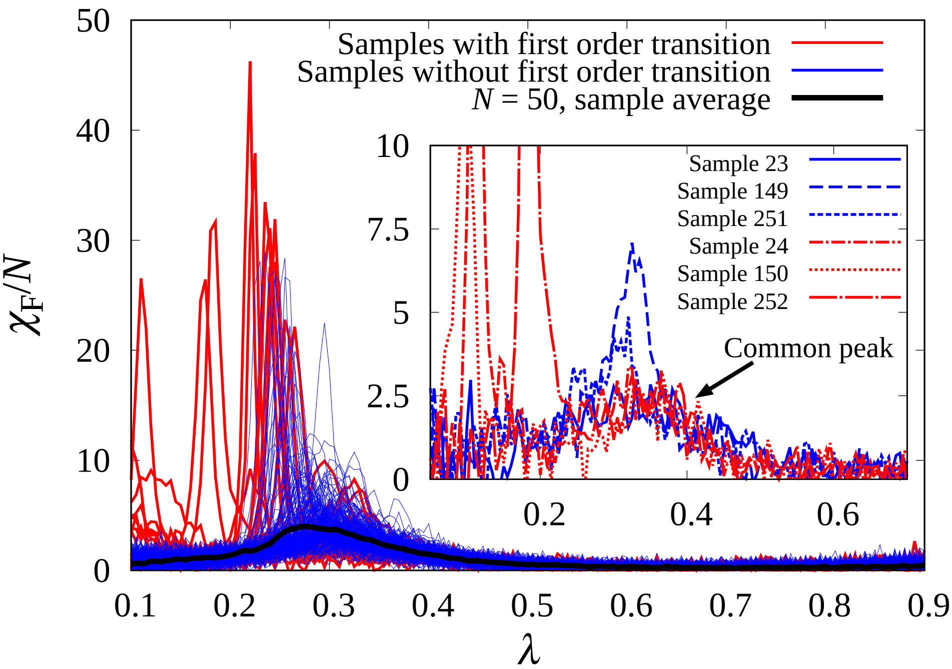

Fig. 6.4: λ dependence of the fidelity susceptibility χ_{F} for 256 samples with $N = 50$. The use of colors is the same as Fig. 6.1. The inset shows a magnified view of 6 samples for clarity. The arrow is pointing at the moderate peak of χ_{F} at $\lambda \sim 0.3$

So far, we have seen that both of spin-glass phase transition scenario and Anderson localization scenario seems to fail for our model of MIS with USE. Then is there no

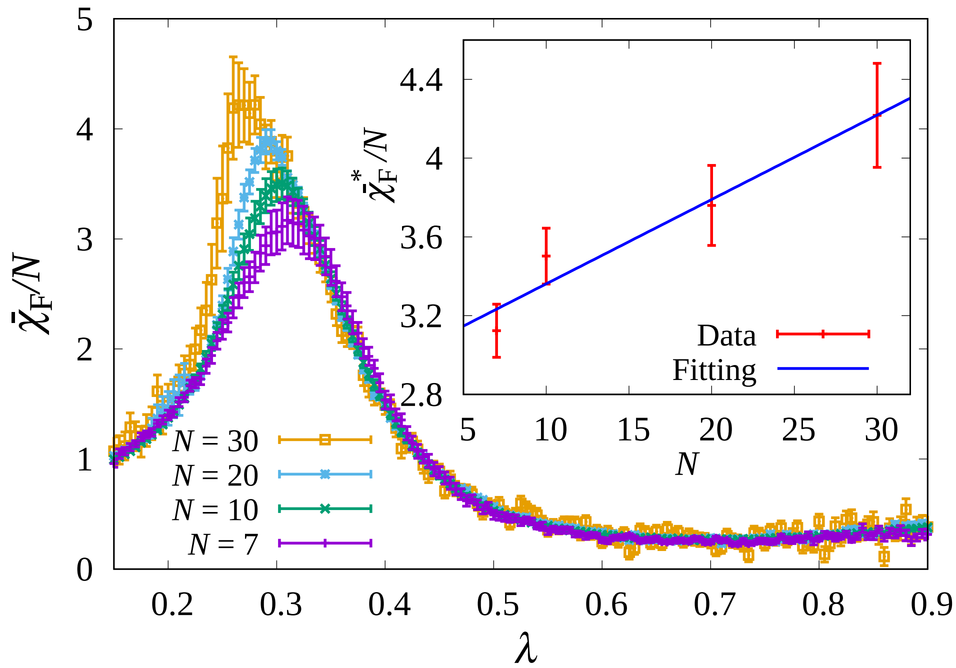

Fig. 6.5: The sample average of normalized fidelity susceptibility $\overline{\chi_F}/N$ is drawn fo r different system sizes. The inset shows peak values and a least square fitting.

"underlying phase transition" at all which enables us to understand the first order phase transition as a phenomenon within a particular phase? Are the first order transitions merely a random phenomena occurring for random instances of NP-hard problems, just to satisfy the NP hardness assumption? It turns out that there *is* an underlying phase transition, which we can detect from the fidelity susceptibility.

Fig. 6.4 presents λ dependence of the fidelity susceptibility χ_F for each sample with $N = 50$. Similarly to q and F_{ans}, χ_F also shows an acute peak at the points where first order transitions occur (See Fig. 6.4). Other than the acute peaks corresponding to the first order transitions, we can see a relatively moderate peak at $\lambda \sim 0.3$. To see this more clearly, the sample average of the fidelity susceptibility $\overline{\chi_F}$ is taken in Fig. 6.5. It shows a diverging trend at $\lambda \sim 0.3$, towards the thermodynamic limit.

One subtlety should be noted here, i.e., although we have seen first order transitions with arguably exponentially small energy gaps, we only have linear inverse temperature scaling as $\beta = 3.5N$. An exponentially large β is needed if ground state properties right at the exact first order transition points are being focused. However, an exponential scaling in β is not only practically impossible for large system sizes which we investigate in our study, but also unnecessary for all points other than the first order transitions with larger energy gaps. There are two observations which suggest that the phase transition associated with the divergence of $\overline{\chi_F}$ only accompanies a polynomially small energy gap $\Delta E(\lambda \sim 0.3)$. One is that the divergence of $\overline{\chi_F}$ is considerably weaker compared to the individual peaks of χ_F at the first order transitions. Another is that $\overline{\chi_F}$ saturates well before $\beta = 3.5N$. Thus, we can be confident that the divergence of $\overline{\chi_F}$ at $\lambda \sim 0.3$ is indeed a property of the ground state.

Importantly, the moderate peak of χ_F in each sample is present in samples of both types, either with or without first order transitions, as seen in the inset of Fig. 6.4. This means that in the thermodynamic limit, all samples should have diverging χ_F at $\lambda \sim 0.3$.

Moreover, all of the first order transitions occur within the low λ phase of this divergence. This implies that there is a phase transition common among all the samples, and the phase is responsible for inducing first order transitions within the phase, compatible with mechanisms discussed in previous studies [101, 150, 151, 152]. However, as we mentioned at the end of the last section, since the sample-averaged $\overline{q}$ does not show any singularity, a simple spin-glass scenario does not explain this phase transition. Nor does the simple Anderson localization scenario hold, since no localization in the configuration space as expected from the previous study was observed.

6.4 Results on the Erdös-Rényi Ensemble

We have seen that the unique solution ensemble has an underlying phase transition which is common among all samples regardless if they are easy/hard instances. The nature of the underlying phase transition however, remained unclear since it did not immediately fit either proposed scenarios of spin glass or Anderson localization. One strategy for further investigating the underlying phase is to consider a similar ensemble which some part of the phase diagram is already *known*, and compare it to the USE. This is exactly what we do in this section.

As we have discussed in section 6.2.2, making the solution unique is quite crucial for fulfilling our motivation to see the physical consequence of the NP-hardness assumption. While the simple Erdös-Rényi ensemble was unsuitable because of the multiple solutions, we know from chapter 5, that the classical MIS problem is in the RSB phase for average degree $c > e = 2.718\ldots$ which was one guideline on how we constructed USE. The classical model corresponds to $\lambda = 0$ (or equivalently $\Gamma = 0$, as we explain below), which means that there should at least be *one* spin-glass phase transition for the QA of MIS with ERE. If that phase transition occurs similarly to the underlying phase transition observed in USE, we can suspect that the underlying transition could be understood as a *modified* spin-glass transition. To sum up, we can further check the validity of the spin-glass picture by measuring the (sample averaged) fidelity susceptibility $\overline{\chi_{\mathrm{F}}}$ for the ERE.

6.4.1 Fidelity Susceptibility χ_{F}

We can clearly see that the sample averaged fidelity susceptibility $\overline{\chi_{\mathrm{F}}}/N$ shows *two* peaks, indicating that there are two phase transitions. Note that here, we are using the Hamiltonian in the form of

$$\hat{H}(\Gamma) = \hat{H}_{\mathrm{P}} + \Gamma \hat{H}_{\mathrm{T}} \tag{6.25}$$

which is essentially the same as the QA Hamiltonian Eq.(6.1) with the transformation of $\Gamma \leftrightarrow \lambda/(1-\lambda)$. Therefore, the moderate peak on the right side around $\Gamma \sim 0.4$ corresponds to $\lambda \sim 0.3$, where the moderate peak for USE occurred.

We further provide Fig. 6.7 to see the tendency towards the thermodynamic limit, on the peaks. The peak of $\overline{\chi}_{\mathrm{F}}$ on the left (around $\lambda \sim 0.1$) may appear to be shifting towards the left. However, we can see from Fig. 6.7 (left) that the least-squares fitting indicates that the divergence of $\overline{\chi}_{\mathrm{F}}$ in the thermodynamic limit will take place in finite Γ. Moreover, it will not be much of a problem even if the phase transition in the left takes place at $\Gamma = 0$ in the thermodynamic limit, since that will only imply that the classical spin-glass phase which we argued in chapter 5 only appears in the classical limit $\Gamma \to 0$.

Also, the peak of $\overline{\chi}_{\mathrm{F}}$ on the left (around $\lambda \sim 0.4$) may seem to be too weak as a divergence, but we do have some support on the divergence as shown in Fig. 6.7 (right). We can see that there is a diverging trend similar to that of the underlying transition for USE (the slope for USE was about 0.04 whereas we have 0.02 for ERE.).

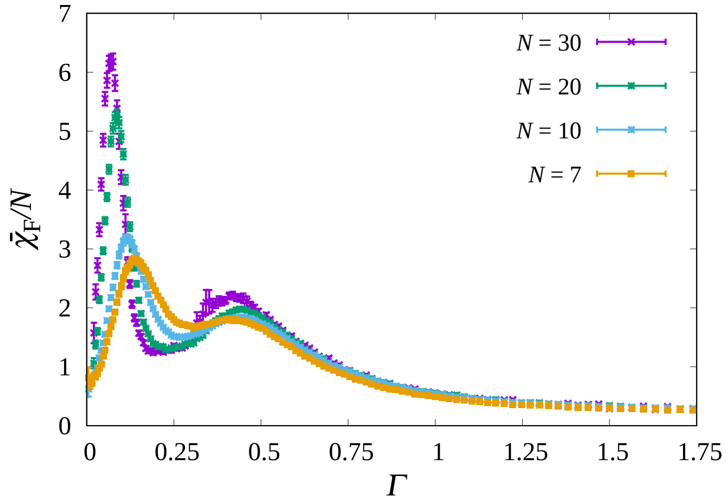

Fig. 6.6: The Γ dependence of sample averaged (normalized) fidelity susceptibility $\overline{\chi_F}(\Gamma)/N$ is drawn for different system sizes.

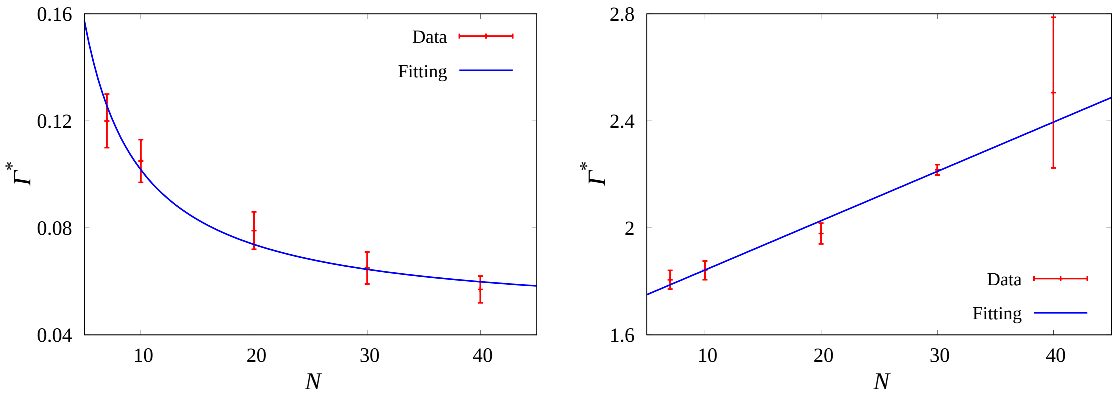

Fig. 6.7: The graph on the left shows the size dependence of where the peak in $\overline{\chi}_F(\Gamma)$ on the left is. A least-squares fitting to the form $\Gamma^*(N) = a + b/N$ is done, giving $a = 0.046(4)$ and $b = 0.55(6)$. The graph on the right shows the size dependence of the peak value in $\overline{\chi}_F(\Gamma)$ on the right. A least-square fitting to the form $\overline{\chi}_F^* = aN + b$ is done, giving $a = 0.018(1)$ and $b = 1.65(2)$.

The fact that there are not one but *two* diverging points for $\overline{\chi}_F$ implies that there are three phases in total. We already know that the phase in the leftmost ($\Gamma \lesssim 0.1$) is the RSB phase as in classical MIS as we discussed in chapter 5. The rightmost phase ($\Gamma \gtrsim 0.4$) is also obviously the quantum paramagnetic phase just like any other stoquastic system. We are left with the unknown phase in between. It is possible that while the leftmost phase corresponds to a classical spin glass where as the intermediate phase could be interpreted as a quantum spin-glass phase. To test this idea, we measure the spin-glass susceptibility.

6.4.2 Spin-glass Susceptibility χ_{SG}

The spin-glass susceptibility χ_{SG} we have introduced in section 5.2.1 was a purely classical quantity which quantified the broadness of $P(q)$. The classical spin-glass susceptibility was defined in Eq.(5.15) as $\chi_{\rm SG} = \frac{1}{N}\sum_{i,j} C_{ij}^2$, where C_{ij} was the correlation between the i th and j th spins, which could be expressed as

$$C_{ij} = \left.\frac{\partial\langle\sigma_i\rangle}{\partial h_j}\right|_{h_j=0} = \beta\left(\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle\right). \tag{6.26}$$

If we interpret this as a quantum system, we get the *canonical correlation* as

$$C_{ij} = \left.\frac{\partial\langle\hat\sigma_i^z\rangle}{\partial h_j}\right|_{h_j=0} = \int_0^\beta \langle\hat\sigma_i^z(\tau)\hat\sigma_j^z\rangle d\tau, \tag{6.27}$$

where $\hat A(\tau) := e^{\tau\hat H}\hat A e^{-\tau\hat H}$ is the imaginary time evolution of $\hat A$. With a spectral decomposition, we obtain

$$\begin{aligned} C_{ij} &= \int_0^\beta \frac{1}{Z}\sum_{n,m}\langle\Psi_n|e^{\tau\hat H}\hat\sigma_i^z|\Psi_m\rangle\langle\Psi_m|e^{-\tau\hat H}\hat\sigma_j^z e^{-\beta\hat H}|\Psi_n\rangle d\tau & (6.28)\\ &= \frac{1}{Z}\sum_{n,m}\langle\Psi_n|\hat\sigma_i^z|\Psi_m\rangle\langle\Psi_m|\hat\sigma_j^z|\Psi_n\rangle\int_0^\beta e^{-(\beta-\tau)E_n}e^{-\tau E_m}d\tau & (6.29)\\ &\xrightarrow{\beta\to\infty} 2\sum_{n\neq 0}\frac{\langle\Psi_0|\hat\sigma_i^z|\Psi_n\rangle\langle\Psi_n|\hat\sigma_j^z|\Psi_0\rangle}{E_n - E_0}. & (6.30)\end{aligned}$$

This gives us the "purely quantum" spin-glass susceptibility*8

$$\chi_{\rm SG}^{(\rm q)} = \frac{4}{N}\sum_{i,j}\left(\sum_{n\neq 0}\frac{\langle\Psi_0|\hat\sigma_i^z|\Psi_n\rangle\langle\Psi_n|\hat\sigma_j^z|\Psi_0\rangle}{E_n - E_0}\right)^2 = \frac{4}{N}\sum_{n,m\neq 0}\frac{\left|\sum_i\langle\Psi_0|\hat\sigma_i^z|\Psi_n\rangle\langle\Psi_m|\hat\sigma_i^z|\Psi_0\rangle\right|^2}{(E_n-E_0)(E_m-E_0)}, \tag{6.31}$$

which somewhat resembles to the spectral representation of the fidelity susceptibility

$$\chi_{\rm F} = \sum_{n\neq 0}\left|\frac{\langle\Psi_n|\frac{\partial}{\partial t}\hat H|\Psi_0\rangle}{E_n - E_0}\right|^2. \tag{6.32}$$

However, currently we have no methods to measure this purely quantum spin-glass susceptibility other than exact diagonalization, and we thus need to consider a more simple way of defining the spin-glass susceptibility in the quantum case.

Since we were measuring the overlap parameter q, and the classical spin-glass susceptibility could be interpreted as the kurtosis of $P(q)$, defining the quantum version analogously is a natural strategy. The actual quantity we were measuring in section 6.3.1 was defined in Eq.(6.6) as

$$\begin{aligned}\langle q\rangle &:= \frac{1}{N}\sum_i\langle\hat\sigma_i^z\rangle^2 & (6.33)\\ &= \langle\frac{1}{N}\sum_i\hat\sigma_i^{z(1)}\hat\sigma_i^{z(2)}\rangle & (6.34)\end{aligned}$$

*8 This is somewhat different from the quantity studied in [170], which is also called the quantum spin-glass susceptibility $\chi_{\rm sg}^{(\rm q)} = \frac{4}{N}\left(\sum_{n\neq 0}\frac{|\langle\Psi_0|\hat\sigma^z|\Psi_n\rangle|^2}{E_n - E_0}\right)^2$.

$$= \frac{1}{N}\sum_i \langle \frac{1}{\beta}\int_0^\beta \hat{\sigma}_i^z(\tau)d\tau\rangle^2 \tag{6.35}$$

$$= \langle \frac{1}{N\beta^2}\sum_i \int_0^\beta \int_0^\beta \hat{\sigma}_i^{z(1)}(\tau_1)\hat{\sigma}_i^{z(2)}(\tau_2)d\tau_1 d\tau_2\rangle. \tag{6.36}$$

While The definition of $\langle q\rangle$ was in the form of Eq.(6.33) or Eq.(6.34), in order to gain high precision for measuring $\langle \hat{\sigma}_i^z\rangle$, we rewrote $\langle q\rangle$ as in Eq.(6.35). This corresponds to not just sampling the basis states $|\boldsymbol{\sigma}\rangle$ for calculating $\langle \hat{\sigma}_i^z\rangle$, but also sample from all L configurations along the Trotter direction[*9].

Therefore, there is subtlety in what we should consider as the q (not $\langle q\rangle$) being sampled. Note that this subtlety only arises in quantum systems. While interpreting $\frac{1}{N}\sum_i \hat{\sigma}_i^{z(1)}\hat{\sigma}_i^{z(2)}$ as q is straightforward, numerically it can only take N different values, which seems too small since we are only dealing with systems with sizes up to $N = 40$ now. On the other hand, interpreting $\frac{1}{N\beta^2}\sum_i \int_0^\beta \int_0^\beta \hat{\sigma}_i^{z(1)}(\tau_1)\hat{\sigma}_i^{z(2)}(\tau_2)d\tau_1 d\tau_2$ as q accordingly to Eq.(6.36) gives a very smooth $P(q)$, but it takes $O(NL^2)$ time to compute, which is practically troublesome. Thus, we take the strategy of considering

$$\frac{1}{N\beta}\sum_i \hat{\sigma}_i^{z(1)}\int_0^\beta \hat{\sigma}_i^{z(2)}(\tau)d\tau \tag{6.37}$$

as q. This quantity is in between the first two, and is computable in $O(NL)$ time while the corresponding $P(q)$ is smooth[*10].

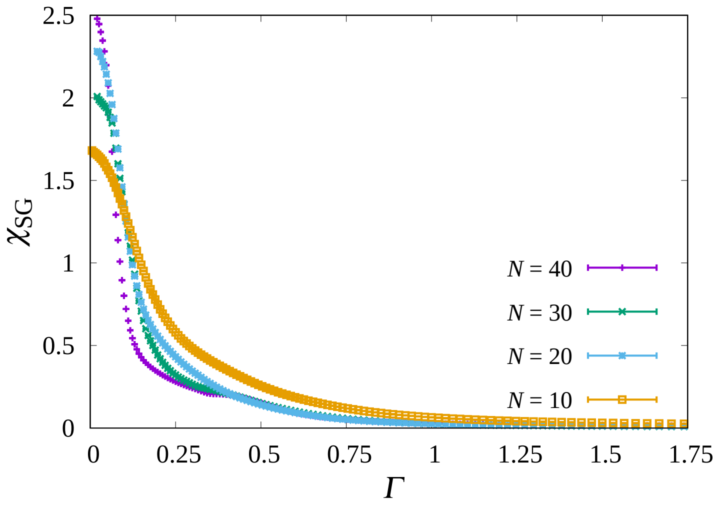

Fig. 6.8: The Γ dependence of sample averaged (normalized) fidelity susceptibility $\overline{\chi_{\mathrm{F}}}(\Gamma)/N$ is drawn for different system sizes.

We can see that the resulting χ_{SG} exhibits only *one* divergence around $\Gamma \sim 0.1$ which corresponds to the divergence of χ_{F} on the left (Fig. 6.8). While there is no divergence

---

[*9] This could be understood from the periodic boundary condition in the Trotter direction.

[*10] It may be arguable that the integrated quantities are "less physical" since they could not be measured directly in an experiment by a projective measurement. While taking the integration may smooth out information (we confirm that is not the case in our model), it is known that the integrated values are experimentally measurable [171].

around $\Gamma \sim 0.4$, there may be a singularity, which further examination is necessary for confirmation.

The important point is that we only see a divergence of χ_{SG} for the phase transition at the leftmost, implying that the phase transition in the right is indeed *not* a spin-glass transition.

6.4.3 Other Quantities

We measured other quantities as well to see if there are any singularities regarding the phase transition in the right side. However, no singularities are found so far, except for the fidelity susceptibility, as it is clear from the figures below.

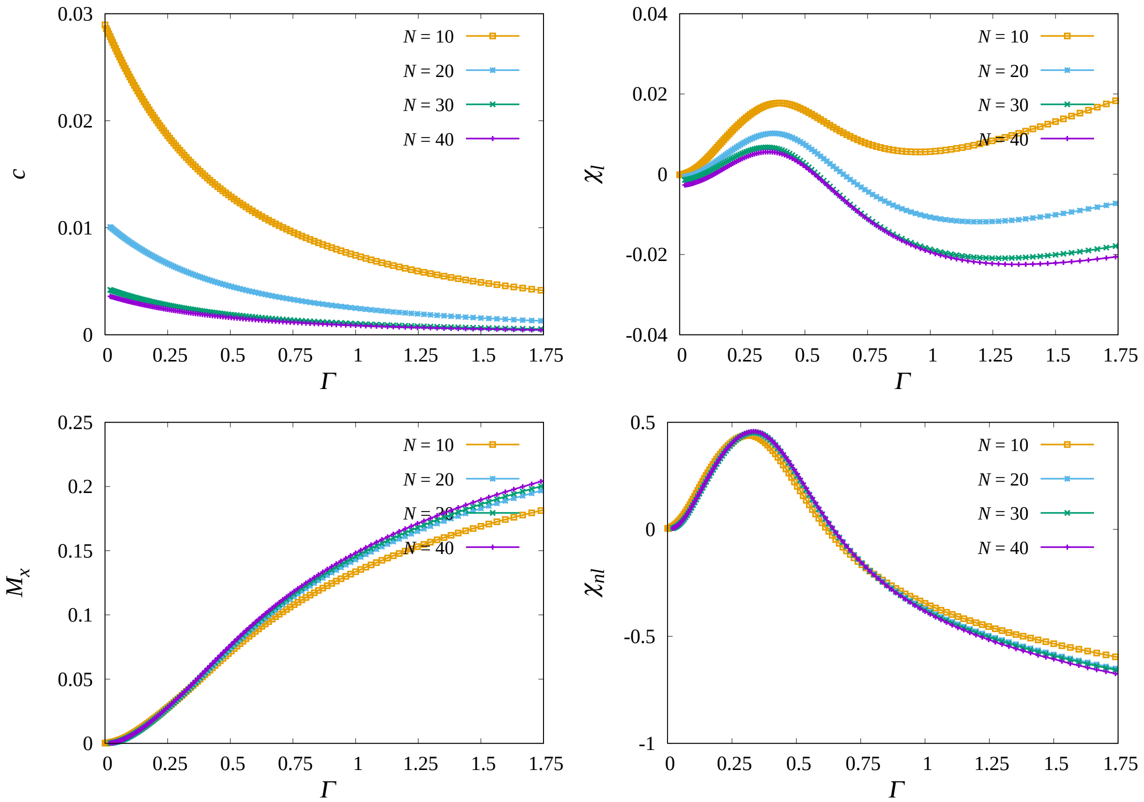

Fig. 6.9: Graphs of the $\Gamma = \lambda/(1-\lambda)$ dependence of sample averaged quantities. The quantities presented are specific heat (left top), transverse magnetization (left bottom), linear magnetic susceptibility (right top), and non-linear magnetic susceptibility (right bottom). All of the quantities seem to converge as the system size grows, and have no singularities.

It should be noted that the non-linear susceptibility is only known to match the spin-glass susceptibility when there is symmetry in the random interactions J_{ij} (see footnote *7 in chapter 5). Since MIS does not have this symmetry, the non-linear susceptibility has different behavior with the spin-glass susceptibility.

We also see no singularity in the answer fidelity, or the second Rényi entropy. These properties are common with the underlying phase transition found in USE, suggesting that the phases have similar properties.

6.5 Conclusion and Discussions

We have demonstrated that the fidelity susceptibility can be useful to find quantum phase transitions which would otherwise have been hard to confirm. For the MIS problem with unique solutions, we find a divergence in $\overline{\chi_{\mathrm{F}}}$ around $\lambda \sim 0.3$, which other conventional quantities fail to capture. The non-singularities in $\overline{q}$ and S_{ans} suggest that a simple spin-glass [172] or a localization picture fails to explain the phase. However, weak (non-RSB) spin-glass transitions or complicated many-body localization are not completely ruled out. Freezing phenomena found in the actual quantum annealing machine [174] may also be connected with the new phase in our result. Many-body localization is also a possibility for the underlying phase transition which is becoming under interest recently [173].

Unfortunately, it is unclear what kind of the phase transition is, since conventional quantities such as the specific heat, transverse magnetization, linear magnetic susceptibility, and the non-linear magnetic susceptibility did not exhibit any singularities as well as more model specific quantities such as F_{ans}. However, it is always possible that a weak singularity is smoothed out due to finite size effects. Thus, to be more precise, we should say that the fidelity susceptibility χ_{F}, compared to other physical quantities, provides an easier way to confirm the existence of a subtle quantum phase transition, as in this case. Previous results [163] show that χ_{F} could be used as indicators for various transitions including spin liquids etc. The present work shows that it is also true for the case of QA of NP-hard problems which is glassy and has frustrated quenched disorders.

Furthermore, the phase transition found by $\overline{\chi_{\mathrm{F}}}$ has important implications for how QA fails to efficiently compute NP-hard problems, regarding the NP hardness assumption. The first order transitions in terms of q found in our study, should accompany exponentially small energy gaps according to arguments made in [71], resulting in exponential running time of QA. These first order transitions are only observed in finite proportion of the samples, also differing in transition points among samples. This is consistent with the fact that NP-hardness is a notion for the worst-case analysis, meaning that not *all* of the samples studied here need to be hard instances. However, notably, the first order transitions occur only in the low-λ side of the $\overline{\chi_{\mathrm{F}}}$ divergence. Remarkably, even the samples with first order transitions have peaks in χ_{F} corresponding to the divergence in $\overline{\chi_{\mathrm{F}}}$ around $\lambda \sim 0.3$. This strongly suggests that the underlying quantum phase captured by $\overline{\chi_{\mathrm{F}}}$ is common among all samples, and is inducing the first order transitions within the phase. While this is partially compatible with proposed scenarios with an underlying phase inducing exponentially small energy gaps [151, 152, 101, 150], they do not fully explain our results as we mentioned above. The proportions of the samples with first order transitions in unique solution NP-hard problems are only investigated numerically so far [71]. Therefore, it should be crucial to identify the nature of this underlying phase transition detected by $\overline{\chi_{\mathrm{F}}}$, in order to fully understand the physical cause leading to the failure of QA for NP-hard problems, and the physical consequences of the NP hardness assumption, or more specifically NP$\not\subseteq$BstoqP in this case.

The fact that a similar phase transition occurs in the ERE as well as USE, suggests that it may be possible to tackle the nature of the underlying phase from replica calculations. Since there are only replica calculations for quantum systems with fully coupled systems currently (such as the quantum SK model [175] or the QREM), approaches applicable to sparse quantum mean-field systems (such as quantum belief propagation [147]) should be adopted.

We finally note that χ_{F} also shows a very sharp peak at the first order transitions which occur within the low λ side of the phase transition. Together with the fact that in some situations χ_{F} could upper bound the size of energy gaps [176, 177], this suggests the possibility that χ_{F} may actually detect all of the transitions which are relevant for QA.

6.A The Unique Solution Ensemble

Instead of using simple Erdös-Rényi random graphs which have multiple solutions for the MIS, we use random graphs which have unique solutions. This ensures that it is always the minimum energy gap $\Delta E_{\min}$ that causes the failure of QA. If there are multiple solutions, a small $\Delta E_{\min}$ would not necessarily imply failure, since it can still end in the degenerate ground state of $\hat{H}_{\mathrm{P}}$.

To generate random-graph ensemble with a unique solution, we randomly add edges to the Erdös-Rényi random graph in the following way. If the original Erdös-Rényi random graph already has a unique solution, we can just use it, although the probability of such a graph occurring decreases as N increases. When the graph has multiple solutions, it means the vertices could be divided into two groups, namely the backbones and the non-backbones. The backbones are the vertices which are constantly in the independent set or out of it throughout all the possible solutions. If a given graph has a unique solution, all the vertices belong to the backbone by definition. After checking which vertices the backbones are, we randomly assign one of the possible solutions, at random. If there are more than two non-backbones which are inside of the assigned solution, we add an edge between those two vertices. This makes the chosen solution no longer valid, while making sure that there are still solutions of the same size.

We continue this process until there are no more pairs of non-backbones inside a particular maximum independent set solution. If there still remains a non-backbone vertex with no pair, we randomly choose one backbone vertex within the solution and add an edge with that. This procedure always decreases the degeneracy. When the degeneracy is totally removed and the solution is unique, the procedure ends successfully. If feasible solutions vanish during this procedure, we discard the graph and start all over again. We call this stochastically generated ensemble of graphs, the unique solution ensemble (USE) in this chapter.

6.B The Dynamic Programming Leaf Removal Algorithm

We should confirm that the process of making the solution unique does not make the problem easier, since we want to know the physical picture of *hard* problems. A specific algorithm which we call the Dynamic Programming Leaf Removal algorithm (DPLR) explained in section 5.3.2 was used for confirming this point.

As we saw in chapter 5, the DPLR algorithm is an algorithm which correctly presents the RS-RSB/easy-hard correspondence in the MIS problem. It thus becomes a good measure of if we are truly considering an average-case hard ensemble or not. Furthermore, the DPLR algorithm is the analogue of the well-known DPLL algorithm [128] for the satisfiability problem. As it is known that the running time of DPLL scales polynomially in the easy parameter region of SAT [178], DPLR should be a fairly good algorithm to see the hardness of the MIS problem.

We can compare the simple Erdös-Rényi random graphs and the unique solution ensemble in terms of computation time using DPLR, from Fig. 6.10 and Fig. 6.11. Since the distribution of time steps has exponentially long tails, it is convenient to focus on the logarithm of the time steps needed. We plot the logarithm of the median time steps needed for finding and confirming a solution for the MIS problem. We used over 1000 samples for each size to estimate the median value, and the error bars are drawn by the

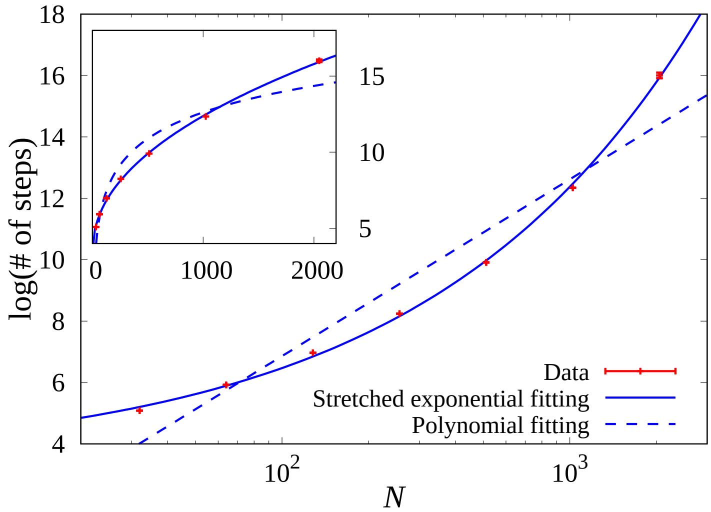

Fig. 6.10: The logarithm of median time steps needed to find the MIS for Erdös-Rényi random graphs using the DPLR algorithm is shown as a function of the system size N. The inset shows the same plot with non-logarithmic scale for the horizontal axis. All lines are obtained by least square fittings, either assuming $y = aN^b + c$ (stretched exponential) or $y = c + d\log(N)$ (polynomial). The stretched exponent b is estimated as $b = 0.465(25)$.

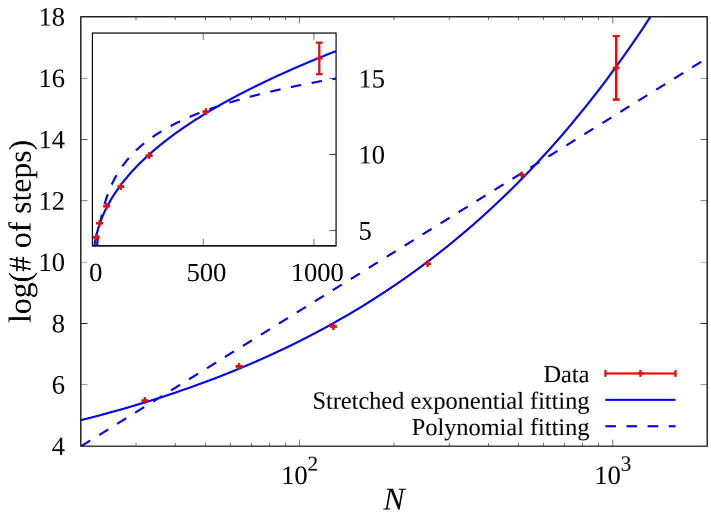

Fig. 6.11: The logarithm of median time steps needed to find the MIS for randomly generated unique solution graphs, using the DPLR algorithm. The inset shows the same plot with non-logarithmic scale for the horizontal axis. The fittings are performed as the same as in Fig. 6.10, yielding that the stretched exponent $b = 0.433(19)$.

bootstrap method. Calculating by the logarithm of the time step allows us to have small error bars, which otherwise would require an exponential amount of data to have constant size error bars. Importantly, the median time step shows stretched exponential scaling for both ensembles. We can see that a polynomial scaling, shown by the dotted lines in Fig. 6.10 and Fig. 6.11, does not fit, and a stretched exponential scaling does. The fact that the median computation time grows stretched exponentially implies that at least half of the samples will be hard instances. The exponents are $N^{0.465\pm0.025}$ and $N^{0.433\pm0.019}$ for Erdös-Rényi random graphs and the unique solution ensemble respectively. The fact that both ensembles scale stretched exponentially and not polynomially implies that the MIS problem is hard for both ensembles and we have not changed the hardness of the problem drastically by making the solution unique described in Appendix 6.A.

Chapter 7

Concluding Remarks and Future Directions

We have seen three studies on topics concerning the relation between physics and computational complexity. Here, we will go through and review what we have shown and their implications, together with possible future directions in research.

In chapter 2, we reviewed the basics of computational complexity theory, which lead to notions such as NP-hardness/completeness, average case complexity, and quantum computation.

We showed reasons for regarding NP-hardness as an archetypical property of "hardness" in general and further introduced the idea that no physical phenomena should be able to efficiently compute such hard problems. This was called the NP hardness assumption. Average case complexity was introduced since the worst case analysis framework may be regarded as too theoretical. However, we reviewed that the average case complexity also faces difficulties, and saw that statistical physics analysis may provide an alternative approach for average case complexity. This perspective was further examined in chapter 5, and then was also applied in chapter 6. Quantum computation was introduced for considering quantum effects on computation. While it could be seen as that allowing probabilistic procedures for Turing machines already introduced probabilistic "superpositions", the coherence in the superposition for quantum states seemed to allow more powerful computations in the class BQP. The boundary of BPP and BQP was further discussed in chapter 4, and then was also applied in chapter 6.

In chapter 3, we reviewed the basics of quantum annealing (QA) and the adiabatic quantum computation (AQC) which are essentially the same but have different emphasis.

AQC was a protocol which was aimed to be a computational protocol from the beginning, and was thus defined with the help of the quantum adiabatic theorem. QA, on the other hand, was originally proposed as a quantum analogue of simulated annealing (SA), and was therefore defined more procedurally. It is possible to consider the originally proposed QA as a restricted version of AQC. However, recent studies of QA are not necessarily restricted as in the originally proposed QA in the sense that they may consider (1) non-adiabatic finite time annealing, (2) noise or finite temperature effects, (3) non-stoquastic terms in Hamiltonian, etc. While the third direction makes QA closer to the idea of AQC by becoming BQP-complete, the first two extensions are rather directed away from AQC. We organized this situation, and explained that we will focus on stoquastic QA in the following chapters (4 and 6), with adiabatic operations.

In chapter 4, we further examined the stoquasticity of the original QA. We explained quantum Monte Carlo (QMC) methods in detail, because they play a crucial role both in understanding stoquasticity and in our study presented in chapter 6.

The path integral Monte Carlo (PIMC) method, a very basic QMC method, was explained first, and then we explained stochastic series expansion (SSE), the main tool in chapter 6. We also reviewed a somewhat different type of Monte Carlo algorithm called diffusion Monte Carlo (DMC). These three methods have different strategies for correctly simulating a quantum system, but remarkably seems to have similar constraints on the quantum Hamiltonian, which is stoquasticity (i.e. absence of the negative sign problem).

While all of the methods allow us to map a stoquastic system to a classical system, the equilibration time required for the Monte Carlo methods to correctly sample the system has been an open problem in general. After reviewing few examples where straightforward QMC or DMC fails to simulate specifically designed stoquastic QA, we focused on simulating stoquastic QA by the QMC, together with the exchange Monte Carlo (EMC, also known as parallel tempering/replica exchange) method, since it is the most used algorithm for simulating QA.

We first showed numerically that the obstruction which was proposed to impede does not affect the behavior of QMC methods, and is thus not an essential obstruction on its own. The focus then, was if the energy landscape of the type proposed for the obstruction for DMC was able to become an obstruction even for QMC methods.

Our result was somewhat surprising in the sense that we found a simple example by extending the obstruction for DMC, where QA is possible to go through in polynomial time (guaranteed by the polynomially large energy gap), but the corresponding simulation using QMC and EMC takes exponential time. The intuitive understanding of EMC is to exchange the replicas along a certain curve in the parameter space to enhance equilibration in the "hard region" where the correct distribution is a priori hard to realize. Our example *does* have a continuous variable λ connecting the hard region (the golf-hole potential) and the easy region (the golf-hole potential with a longitudinal magnetic field as a hint). Although EMC achieves high exchange rate in our example, they require exponentially long time for finding the golf-hole potential for parameter region close to $\lambda = 1$.

The first implication of our result is that there is unaddressed quantum weakness of EMC, perhaps a weakness similar to a previously addressed classical weakness. While this does not immediately imply inadequateness of EMC for the usage of simulating QA as done in many researches since they have a qualitatively different structure of the energy landscape, it is interesting to consider a naturally arising model where a golf-hole like potential emerges.

The second implication of our result is that EMC is not sufficiently strong enough in order to be utilized for proving claims such as BstoqP=BPP. While separation of complexity classes require extremely non-trivial proof in general, equivalences are often proved constructively through clever algorithms. Our example shows that the EMC in its basic form cannot be that "clever algorithm". We easily see numerically that our example is easily simulable by simulated quantum annealing (SQA), which is currently the remaining algorithm for proving equivalences. Since there are already proposed examples where naive SQA fails from topological obstructions, a proof of BstoqP=BPP should require similar techniques as population annealing, or at least repeating SQA polynomially many times. It is interesting that both of the possible algorithms are dynamic, in the sense that they have time-evolving Hamiltonians defining a stochastic dynamics non-uniform in time. The importance of this point remains unclear.

There are several possible directions for further research on this topic. While proving BstoqP=BPP is an obvious goal, numerical approaches for supporting that may come first. For example, numerically showing that population annealing or repeated SQA can efficiently simulate stoquastic QA with the topological obstructions discussed in section 4.3.1 would be interesting. This is because while for the first few examples provided in [106] were evidently simulable by polynomially repeated SQA as we discussed in section 4.3.1, it is nontrivial for the examples provided later (which we did not discuss). It would be worth confirming that if a simple repetition of SQA will be enough, or resampling methods such as population annealing will be necessary.

It is also interesting to consider a classical Hamiltonian having the same property as our example of having an exponentially long equilibration time while having a high enough exchange rate. This seems possible since mapping our stoquastic Hamiltonian back to a classical one has no difficulty in principle. In that case, we may be able to compare the strengths and weakness of SA and EMC, just as comparing SQA and QMC+EMC, further investigating the difference between classical probability and stoquastic probability.

In chapter 5, we focused on the connection between average case complexity and statistical physics from in the classical setting. We constructed a simple novel algorithm DPLR which exhibits precise correspondence between the average case complexity (lower bounded by DPLR) and statistical mechanical phases.

We focused on the maximum independent set (MIS) problem, as a simple NP-complete/hard problem. This problem enables us to consider random graphs as the random instances. We first reviewed that from the NP hardness assumption perspective, whether the setting we consider is NP-hard or NP-complete does not matter because P=NP⇒P=PH.

We then reviewed the situation of the statistical mechanical approach to random NP-hard problems, after reformulating the MIS problem as a variation of the Ising spin-glass model. The replica symmetry breaking (RSB) phase transition was defined, and its properties were discussed. While the rugged energy landscape arising in the RSB phase intuitively suggests hardness in searching the ground state in the RSB region, we saw that the situation is generally not so simple as in the case for random 3SAT. While the situation is completely open for SAT, we aim for seeking correspondence for the random MIS problem.

We introduced the leaf removal (LR) algorithm, and then the dynamic programming leaf removal (DPLR) algorithm which we propose. We saw that for Erdös-Rényi random graphs, the LR algorithm already displays a correspondence between the typical computation time and the (full) RSB phase transition. However, for scale free networks which we introduced next, this is not the case and LR fails well before the RSB transition point. We found that while DPLR takes typically exponential time in the RSB region, its typical computation time only scales linearly to the problem size in the entire RS region. Finite size effects were observed within the RS phase, with an interesting scaling behavior. The problem size N^* where the typical computation time no longer shows finite size effects increase polynomially as we get closer to the RS/RSB transition point, and finally diverges at the phase transition point. This could be seen as a critical behavior in typical computation time.

There are several implications and directions for further research for our result in this chapter. The first implication is obviously that the RS-RSB/easy-hard correspondence seems to strictly hold in MIS, regardless of the random distribution of instances. It is important that the correspondence did not hold when we only used the LR algorithm. This implies that strong enough algorithms are necessary for searching the connections between statistical physics and average case complexity. The fact that for the MIS problem, such a simple algorithm as DPLR would be enough to exploit the correspondence is interesting. It would be interesting to see the capability of algorithms and also the correspondence, when we use the state-of-the-art algorithm together with backtracking and dynamic programming. This is because both in MIS (DPLR) and SAT [123], this strategy seems to provide the most powerful algorithm so far, working in linear time up to some phase transition point.

The second implication is that the average case complexity jumps from linear time to exponential time, which is rather surprising. This linear to exponential jump is also observed in random 3SAT, and may be a universal feature among random NP-hard problems. While no explanation for this phenomena is known so far, our result at least suggests that it is hard to simply construct a natural nonlinear polynomial time average case complexity class. It would be interesting to consider naturally arising ensembles where the computation time grows polynomially (and not linearly) in some finite parameter region, for two reasons. First, it would be an interesting example on glassy systems where we can explicitly evaluate the equilibration time and its divergence. Second, it may provide us insights on why only low degree polynomials as time complexity classes are far more often observed than polynomials with enormous degrees.

The third implication of our result is that while 1 RSB transitions do not affect the computation time in SAT, full RSB seems to be crucial for computation time for the exact solution. This point has not been discussed so much, since in physics an $O(1)$ difference in energy is usually ignored in the thermodynamic limit. It would be interesting to seek, for example, NP-hard problems which the randomized version only has a 1 RSB phase transition and no full RSB. It is also interesting to consider ensembles which are *natural* for another NP-complete problem. This type of ensemble could be easily made by first generating a random instance simply for one NP-hard problem, and then converting it to another NP-hard problem via the polynomial mapping reductions we discussed.

Equipped with the results from chapter 4 and 5, we analyzed the physical consequence of the NP hardness assumption for an average-case hard problem in the stoquastic setting in chapter 6. The methods such as EMC and SSE introduced in Chapter 4 enablee us to simulate the stoquastic QA, and insights from chapter 5 provided us evidence on average case hardness for the ensemble we used in this chapter.

The NP hardness assumption in this case claims that there should be exponentially small energy gaps for QA computing NP-hard problems (at least for the worst instances). By assuming that the average case hardness for the random ensemble in the RSB phase fully captures the hardness of NP, we can further claim that QA will face typically an exponentially small energy gap for the problem instances from that ensemble.

We first explained two major scenarios for explaining how such exponentially small energy gap may arise. They are the spin-glass scenario and the Anderson localization scenario, both of them not being confirmed in the general case for NP-hard problems. Both of the scenarios present a picture where there is an "underlying phase transition" (either a spin-glass transition or a localization transition), and then *within* the underlying phase, there occur exponentially small energy gaps. Our motivation was to see which of the scenarios is true, or if there exists such an underlying phase in the first place.

After explaining the methods and measured quantities, we first focused on our results on the unique solution ensemble (USE), where the exponentially small energy gaps *must* occur according to the above-mentioned assumptions. We observed that the instances could be divided into two groups, the easy instances and hard instances, distinguishable from the existence of a first order phase transition in q, when dealing with problem sizes up to 40$\sim$50. While these first order transitions could be observed using straightforward quantities such as the spin-glass order parameter q or the answer fidelity F_{ans}, those quantities did not show any singularities indicating "underlying phase transition".

An important finding is that the fidelity susceptibility χ_{F} captures not only the first order transitions but also the underlying phase transition, after we take the sample average $\overline{\chi_{\mathrm{F}}}$. While the existence of the underlying phase transition itself is compatible with both scenarios previously proposed, the measured quantities indicated that the phase transition is neither a normal spin-glass transition or an Anderson localization transition.

We further checked this point by considering the Erdös-Rényi ensemble (ERE) which a spin-glass phase is already known to exist in the classical limit. The expectation is that both of the ERE and USE we used have similar statistical properties, although this point remains to be further analyzed. We found *two* phase transitions using the fidelity susceptibility for ERE. Interestingly, one phase transition occurs at a similar parameter point as for the underlying phase transition in USE. From measuring the spin-glass susceptibility χ_{SG}, we can confirm that the second phase transition which takes place in the parameter region very close to the classical limit is indeed a spin-glass transition. Since χ_{SG} does not diverge at the other phase transition point, it is likely that the underlying phase is *not* a spin-glass phase.

Obviously, figuring out the true nature of this underlying phase transition is one big future direction to pursue. As we mentioned in chapter 6, freezing phenomena in the actual quantum annealer device is reported, which occurs from some value λ similarly to our underlying phase transition. There are also studies on many-body localization (MBL) for quantum spin-glass systems recently. It is possible that the underlying phase transition could be understood as an MBL phenomenon, since it is known that dynamical information of the system could be extracted from SSE samplings. At this point, we also have results that the probability of observing a non-answer maximal independent set after a projective measurement on the z-axis becomes the largest around the underlying phase transition point. It is possible that the underlying phase transition could be understood through the amplitude contributions of these "locally optimal" solutions to the ground state.

Regarding the NP hardness assumption, there are also numbers of interesting future directions of research. It would be interesting to consider a similar situation but with a *totally quantum* AQC, which may lead to an even more quantum phenomena preventing polynomial time computation of the NP-hard problem. While it is numerically difficult to deal with non-stoquastic systems, and it is almost surely impossible to directly observe the first order transitions in that case (since they only occur in large sizes $N \gtrsim 25$), we may be able to see the underlying transitions since they occur from smaller system sizes.

Applying the NP hardness assumption to other areas is also attractive, since it may open up new approaches for complex systems. For instance, a possibility of using computational complexity ideas for proving results on fluid equations has been proposed [179].

We conclude that while we attained results such as the RS-RSB/easy-hard correspondence for the MIS problem, or showing the existence of a novel underlying phase in stoquastic QA for the MIS problem, the attempt to study physics from the computational complexity perspective still has many open problems ahead. We believe that once the basic obstacles such as finding/defining a natural arising average case-complete problem, or showing the exact equivalence/separation of stoquasticity to other classes, the perspective will become extremely fruitful.

謝辞

はじめに指導教官であり、修士課程から計 5 年間私を指導してくださった福島孝治教授には深く深く感謝いたします。僕は何かと問題児だったかもしれないと反省していますが、そんな自分に対し粘り強く指導してくださったことは本当にありがたいです。自分に甘えがあってたるんでいる時にはお叱りを受け、また落ち込みそうな時には激励していただきました。研究に関する議論の際には様々な助言をいただくとともに、研究者としての作法を学びました。他所の研究室で発表をする際や議論をする際の心構え、研究者の世界でどのように生きていくかなど他の様々な面でもご指導いただきました。掛けて頂いた言葉を胸に今後の研究者人生を歩んでいきます。また、そもそも僕が統計力学を知ったのは大学一年生の夏学期に福島さんによる統計力学の授業を受けた時です。それまで僕は名前の響きだけから (?)「統計を使うなんて妥協っぽいイマイチな分野なんだろうな〜」となんとなく思っていのですが、その学期に統計力学の授業を受講して以降はむしろ統計力学こそが一番面白いのではないかとすら思えました。そのような縁に恵まれて今の自分の研究があるのだと不思議な感慨があります。

京都大学の佐々真一教授には大学院時代直接指導を受けることはありませんでしたが、学部 (教養学部基礎科学科) 時代から一貫して議論や激励をして頂きました。感謝しております。今にして思うと、学部三年生の時に佐々さんと一緒に行った Nature of Computation [17] の輪行で僕の興味は大きく方向付けられたと思います。福島さんの統計力学の授業と共に、今の自分の研究を方向づけている輪行セミナーの最終日、ピザ片手に僕が 10 時間くらい話した日は忘れません。今後も研究姿勢や哲学に関して大いに勉強させていただこうと思います。

僕の一年先輩であり、現在名古屋工業大学の助教でいらっしゃる高邉賢史さんには非常にお世話になりました。日常の議論から諸申請書の書き方に到るまで、本当にありがとうございました。また、本博士論文第 5 章は高邉さんとの共同研究になっております。興味を共にし、いつでも議論ができる一年上の先輩がいたことは僕にとって本当に幸せなことでした。感謝申し上げます。

同期の西川宜彦くんにも数え切れないくらいお世話になりました。彼とは物理の議論も思う存分できた上 (僕の研究に対して一番ピントが合っている他人です)、僕がよく大小様々な問題にぶちあたっている時に快く解決を手伝ってくれました。僕も同じくらい彼に対して手を貸せていれば良いのですが (主に英語ではお世話しました)、仮に偏りがあったとしても西川理論によれば「貸しを作れている状態が良い」らしいので、それでも良しということにします (?)。また、非常に良き雑談相手、愚痴相手、ボードゲーム相手でもありました。本当にありがとう！いつも彼は「惇くんはどうせ数年以内に僕の名前を忘れる」などと吹聴していますが、彼の名前は一生絶対に忘れません。

大関真之准教授、田中宗准教授、小渕智之助教、鳩村拓矢くんには様々なセミナーにお呼びして頂き、物理学と情報科学の観点から様々なコメントや激励の言葉をいただきました。特に大関さんには量子アニーリングに関する議論をたくさんして頂くと共に兄貴分として様々な助言を頂き本当に感謝しております。小渕さんの飄々としつつも学問的に深く通じている感じも是非「統計力学飄々としてる勢」の後を継ぐべく今後も見習って行きたいです。

藤井啓祐准教授、金子邦彦教授、清水明教授、沙川貴大准教授には本博士論文の審査員として非常にお世話になりました。特に藤井さんには修士時代から会うたびにアドバイスなどを頂き続けました。教えて頂いた DQC1 や IQP などの below BQP の計算量クラスの存在は本論文の第 4 章の内容に直接繋がりました。また、金子さんには学部の卒業研究においても指導教官としてお世話になり、大学院時代も常に気にかけて頂きました。清水さんには学部時代から授業等でご指導いただくと共に、弓削さんの研究に関する議論を通じて僕の研究に関しても

助言を頂きました。沙川さんと清水さんには修士論文の際にも審査員としてお世話になりました。全員に感謝申し上げます。

設楽智洋くん、松本啓史准教授、白石直人博士、田島裕康博士、Seth Lloyd 教授、樺島祥介教授、越田真史くん、水野雄太くん、高橋昂くん、観山正道助教、星野晋太郎助教、正木祐輔さん、高木隆司くん、村下湧音くん、関優也助教、西口大貴博士、奥山真佳くんをはじめとする諸先同後輩方には、今日に至るまで様々な場面でたくさんの物理の議論だけでなく相談にも乗って頂き感謝申し上げます。特に星野さんには同室だった際にも他所へ移られた後も、研究室や分野が異なるにも関わらず非常に積極的に議論して頂き、更には申請書の書き方なども正確なアドバイスをくださり、真に理想的なロールモデルでした。大変感謝しております。高木くんと設楽くんは高校時代の物理好きな仲良し三人組 (エアーリーダーズ) であり、いつでも物理だけでなくあらゆることを議論・祝福・慰め・相談・楽しむことができました。場所や道は多少違えど三人ともここまで来れて非常に嬉しく思います。観山さんには、申請書や公募での心構えなどをアドバイス頂き、大変感謝しております。正木さんも良き先輩でした。研究ネタに関する議論が盛り上がっただけでなく、様々な非自明共通点もあることが分かり、そのような人と大学院時代を過ごせたことは興味深いです。

中村統太教授、田中篤司助教、田崎晴明教授、森貴司助教、宮崎涼二助教、弓削達郎助教、徳田悟博士、中島千尋博士にも議論をして頂き感謝します。福島研の方々、中西義典助教、小松尚登博士、酒井佑士博士、安倍雅史くん、そして加藤雄介教授をはじめとする加藤研の方々、篠崎美沙子さん、福井毅勇くん、堤康雅博士、寒川崇生くん、武田陸歩くんには日頃の議論や雑談を始め大変お世話になりました。また秘書の長谷眞紀子さんと豊田久子さんには非常にお世話になりました。本当にありがとうございました。

東京大学博士課程教育リーディングプログラム「多文化共生・統合人間学 (IHS) プログラム」からは資金的な援助を頂きました。感謝いたします。高校時代までは「宇宙をやり直したらほぼ確実に歴史や地理は全く違うものになるのに、歴史や地理なんかの非普遍的な学問をやる価値あるの？」なんて狭い見方をしていましたが、IHS プログラムでの多様な活動を通じて、今ある社会問題や歴史は直接的/間接的に自分に関わってくることであり、かつある種の必然性を持って生じているのだということを理解しました。このように視野が広がったことは確実に自分の今後の人生をより豊かに過ごし、そしてより良い決断を下すために必ず役に立つことと思います。

最後に僕を産んでくれ、ここまで育ててくれ、さらには資金的にも援助してくれた両親に感謝いたします。僕が自分の真にやりたいことを見出し、ここまで来る決断ができたのも両親の教育があってのものです。ありがとう！

References

[1] Turing, Alan M. "Computing machinery and intelligence." Mind 59.236 (1950): 433-460.

[2] Deutsch, David. The fabric of reality. Penguin UK, 1998.

[3] Aaronson, Scott. "Why philosophers should care about computational complexity." arXiv preprint arXiv:1108.1791 (2011).

[4] Poincaré, Henri. "Sur le probléme des trois corps et les équations de la dynamique." Acta Mathematica 13. 1 (1890).

[5] Cubitt, Toby S., David Perez-Garcia, and Michael M. Wolf. "Undecidability of the spectral gap." Nature 528.7581 (2015): 207-211.

[6] Moore, Cristopher. "Unpredictability and undecidability in dynamical systems." Physical Review Letters 64.20 (1990): 2354.

[7] Haldane, F. D. M. "Nonlinear field theory of large-spin Heisenberg antiferromagnets: semiclassically quantized solitons of the one-dimensional easy-axis Néel state." Physical Review Letters 50.15 (1983): 1153.

[8] Feynman, Richard P. "Simulating physics with computers." International journal of theoretical physics 21.6 (1982): 467-488.

[9] Deutsch, David. "Quantum theory, the Church-Turing principle and the universal quantum computer." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 400. No. 1818. The Royal Society, 1985.

[10] Calude, Cristian Solin, and Elena Calude. "The Road to Quantum Computational Supremacy." arXiv preprint 1712.01356 (2017).

[11] Turing, Alan Mathison. "On computable numbers, with an application to the Entscheidungsproblem."Proceedings of the London mathematical society 2.1 (1937): 230-265.

[12] Church, Alonzo. "An unsolvable problem of elementary number theory." American journal of mathematics 58.2 (1936): 345-363.

[13] Kleene, Stephen Cole. "General recursive functions of natural numbers." Mathematische annalen 112.1 (1936): 727-742.

[14] Turing, Alan M. "Computability and λ-definability." The Journal of Symbolic Logic 2.4 (1937): 153-163.

[15] Bernstein, Ethan, and Umesh Vazirani. "Quantum complexity theory." SIAM Journal on Computing 26.5 (1997): 1411-1473.

[16] Sipser, Michael. Introduction to the Theory of Computation. Vol. 2. Boston: Thomson Course Technology, 2006.

[17] Moore, Cristopher, and Stephan Mertens. The nature of computation. OUP Oxford, 2011.

[18] Arora, Sanjeev, and Boaz Barak. Computational complexity: a modern approach. Cambridge University Press, 2009.

[19] Schönhage, Arnold. "On the power of random access machines." International Colloquium on Automata, Languages, and Programming. Springer, Berlin, Heidelberg, 1979.

[20] Copeland, B. Jack. "Hypercomputation: philosophical issues." Theoretical Com-

puter Science 317.1-3 (2004): 251-267.
[21] Nielsen, Michael A., and Isaac Chuang. "Quantum computation and quantum information." (2002): 558-559.
[22] Karatsuba, A., and Y. Ofman. "Multiplication of multiple-digit numbers with computers." Doklady Akademii Nauk SSR. Vol. 145. No. 2. 1962.
[23] Schonhage, Arnold, and V. Strassen. "Fast multiplication of large numbers." Computing 7.3-4 (1971): 281.
[24] Papadimitriou, Christos H. "Computational complexity. 1994."
[25] Immerman, Neil. "Languages that capture complexity classes." SIAM Journal on Computing 16.4 (1987): 760-778.
[26] Agrawal, Manindra, Neeraj Kayal, and Nitin Saxena. "PRIMES is in P." Annals of mathematics (2004): 781-793.
[27] Rivest, Ronald L., Adi Shamir, and Leonard Adleman. "A method for obtaining digital signatures and public-key cryptosystems." Communications of the ACM 21.2 (1978): 120-126.
[28] Fortnow, Lance. "The status of the P versus NP problem." Communications of the ACM 52.9 (2009): 78-86.
[29] Cook, Stephen A. "The complexity of theorem-proving procedures." Proceedings of the third annual ACM symposium on Theory of computing. ACM, 1971.
[30] Levin, Leonid Anatolevich. "Universal sequential search problems." Problemy Peredachi Informatsii 9.3 (1973): 115-116.
[31] Aaronson, Scott, and Alex Arkhipov. "The computational complexity of linear optics." Proceedings of the forty-third annual ACM symposium on Theory of computing. ACM, 2011.
[32] Bennett, Charles H., and John Gill. "Relative to a Random Oracle A, $\mathsf{P}^A \neq \mathsf{NP}^A \neq \mathsf{co}\text{-}\mathsf{NP}^A$ with Probability 1." SIAM Journal on Computing 10.1 (1981): 96-113.
[33] Kaye, Phillip, Raymond Laflamme, and Michele Mosca. An introduction to quantum computing. Oxford University Press, 2007.
[34] Aaronson, Scott. "Guest column: NP-complete problems and physical reality." ACM Sigact News 36.1 (2005): 30-52.
[35] Hartmanis, Juris. "Gödel, von Neumann and the P=? NP Problem." Current Trends in Theoretical Computer Science 40 (1993): 445-450.
[36] Levin, Leonid A. "Average case complete problems." SIAM Journal on Computing 15.1 (1986): 285-286.
[37] Bernoulli, Daniel. "Exposition of a new theory on the measurement of risk." The Kelly Capital Growth Investment Criterion: Theory and Practice. 2011. 11-24.
[38] Kirkpatrick, S., and D. Sherrington. "Solvable model of a spin-glass." Phys. Rev. Lett 35 (1975): 1792-1796.
[39] Barahona, Francisco. "On the computational complexity of Ising spin glass models." Journal of Physics A: Mathematical and General 15.10 (1982): 3241.
[40] Gill, John. "Computational complexity of probabilistic Turing machines." SIAM Journal on Computing 6.4 (1977): 675-695.
[41] Nisan, Noam, and Avi Wigderson. "Hardness vs. randomness." Foundations of Computer Science, 1988., 29th Annual Symposium on. IEEE, 1988.
[42] Impagliazzo, Russell, and Avi Wigderson. "P= BPP if E requires exponential circuits: Derandomizing the XOR lemma." Proceedings of the twenty-ninth annual ACM symposium on Theory of computing. ACM, 1997.
[43] Hales, Lisa, and Sean Hallgren. "An improved quantum Fourier transform algorithm and applications." Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on. IEEE, 2000.

[44] Shor, Peter W. “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer.” SIAM review 41.2 (1999): 303-332.

[45] Deutsch, David, and Richard Jozsa. “Rapid solution of problems by quantum computation.” Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 439. No. 1907. The Royal Society, 1992.

[46] Farhi, Edward, et al. “A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem.” Science 292.5516 (2001): 472-475.

[47] Preskill, John. “Quantum computing and the entanglement frontier.” arXiv preprint arXiv:1203.5813 (2012).

[48] Wiesner, Karoline. “The careless use of language in quantum information.” arXiv preprint arXiv:1705.06768 (2017).

[49] Bennett, Charles H., et al. “Strengths and weaknesses of quantum computing.” SIAM journal on Computing 26.5 (1997): 1510-1523.

[50] Aharonov, Dorit, and Itai Arad. “The BQP-hardness of approximating the Jones polynomial.” New Journal of Physics 13.3 (2011): 035019.

[51] Kadowaki, Tadashi, and Hidetoshi Nishimori. “Quantum annealing in the transverse Ising model.” Physical Review E 58.5 (1998): 5355.

[52] Farhi, Edward, et al. “A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem.” Science 292.5516 (2001): 472-475.

[53] Boixo, Sergio and Rønnow, Troels F. and Isakov, Sergei V. and Wang, Zhihui and Wecker, David and Lidar, Daniel A. and Martinis, John M. and Troyer, Matthias. “Evidence for quantum annealing with more than one hundred qubits.” Nature Physics. Vol. 10. (2014).

[54] Messiah, Albert. “Quantum mechanics. Vol. 2 (german translation).” (1979).

[55] Aharonov, Dorit and Wim van Dam, Julia Kempe, Zeph Landau, Seth Lloyd, Oded Regev. “Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation” SIAM Journal of Computing, Vol. 37, Issue 1, p. 166-194 (2007)

[56] Lloyd, Seth. “Universal quantum simulators.” SCIENCE-NEW YORK THEN WASHINGTON- (1996): 1073-1077.

[57] Farhi, Edward, et al. “Quantum computation by adiabatic evolution.” arXiv preprint quant-ph/0001106 (2000).

[58] Aharonov, Dorit, and Michael Ben-Or. “Fault-tolerant quantum computation with constant error.” Proceedings of the twenty-ninth annual ACM symposium on Theory of computing. ACM, 1997.

[59] Shor, Peter W. “Fault-tolerant quantum computation.” Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on. IEEE, 1996.

[60] Vinci, Walter, Tameem Albash, and Daniel A. Lidar. “Nested quantum annealing correction.” npj Quantum Information 2 (2016): 16017.

[61] Albash, Tameem, Victor Martin-Mayor, and Itay Hen. “Temperature scaling law for quantum annealing optimizers.” Phys. Rev. Lett. 119, 110502 (2017).

[62] O ’ Gorman, Bryan, et al. “Bayesian network structure learning using quantum annealing.” The European Physical Journal Special Topics 224.1 (2015): 163-188.

[63] Maddox, John. “Crystals from first principles.” Nature 335.6187 (1988): 201.

[64] Metropolis, Nicholas, et al. “Equation of state calculations by fast computing machines.” The journal of chemical physics 21.6 (1953): 1087-1092.

[65] Hastings, W. Keith. “Monte Carlo sampling methods using Markov chains and their applications.” Biometrika 57.1 (1970): 97-109.

[66] Geman, Stuart, and Donald Geman. “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.” IEEE Transactions on pattern analysis and machine intelligence 6 (1984): 721-741.

[67] Somma, Rolando D., Cristian D. Batista, and Gerardo Ortiz. “Quantum approach

to classical statistical mechanics." Physical review letters 99.3 (2007): 030603.

[68] Tanaka, Atushi, and Manabu Miyamoto. "Quasienergy anholonomy and its application to adiabatic quantum state manipulation." Physical review letters 98.16 (2007): 160407.

[69] Schützhold, Ralf, and Gernot Schaller. "Adiabatic quantum algorithms as quantum phase transitions: First versus second order." Physical Review A 74.6 (2006): 060304.

[70] Laumann, C. R., et al. "Quantum adiabatic algorithm and scaling of gaps at first-order quantum phase transitions." Physical review letters 109.3 (2012): 030502.

[71] Young, A. P., S. Knysh, and V. N. Smelyanskiy. "First-order phase transition in the quantum adiabatic algorithm." Physical review letters 104.2 (2010): 020502.

[72] Seki, Yuya, and Hidetoshi Nishimori. "Quantum annealing with antiferromagnetic fluctuations." Physical Review E 85.5 (2012): 051112.

[73] Wang, Fugao, and D. P. Landau. "Efficient, multiple-range random walk algorithm to calculate the density of states." Physical review letters 86.10 (2001): 2050.

[74] Johnson, Mark W., et al. "Quantum annealing with manufactured spins." Nature 473.7346 (2011): 194-198.

[75] Denchev, Vasil S., et al. "What is the computational value of finite-range tunneling?." Physical Review X 6.3 (2016): 031015.

[76] D-Wave Systems Inc. Homepage. Retrieved December 3, 2017, from https://www.dwavesys.com/our-company/meet-d-wave

[77] Kivlichan, Ian. "On the complexity of stoquastic Hamiltonians." (2015).

[78] Farhi, Edward, and Aram W. Harrow. "Quantum supremacy through the quantum approximate optimization algorithm." arXiv preprint arXiv:1602.07674 (2016).

[79] Cubitt, Toby, and Ashley Montanaro. "Complexity classification of local Hamiltonian problems." SIAM Journal on Computing 45.2 (2016): 268-316.

[80] Landau, L. D. "On the theory of transfer of energy at collisions II." Phys. Z. Sowjetunion 2.46 (1932): 7.

[81] Zener, Clarence. "Non-adiabatic crossing of energy levels." Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 137. No. 833. The Royal Society, 1932.

[82] Somma, Rolando D., Daniel Nagaj, and Mária Kieferová. "Quantum speedup by quantum annealing." Physical review letters 109.5 (2012): 050501.

[83] Amin, Mohammad H. "Searching for quantum speedup in quasistatic quantum annealers." Physical Review A 92.5 (2015): 052323.

[84] Hormozi, Layla, et al. "Nonstoquastic Hamiltonians and quantum annealing of an Ising spin glass." Physical Review B 95.18 (2017): 184416.

[85] Ando, Tsuyoshi. "Totally positive matrices." Linear algebra and its applications 90 (1987): 165-219.

[86] Suzuki, Masuo. "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 (1976): 1454-1469.

[87] Turitsyn, Konstantin S., Michael Chertkov, and Marija Vucelja. "Irreversible Monte Carlo algorithms for efficient sampling." Physica D: Nonlinear Phenomena 240.4 (2011): 410-414.

[88] Suwa, Hidemaro, and Synge Todo. "Markov chain Monte Carlo method without detailed balance." Physical review letters 105.12 (2010): 120603.

[89] Cubitt, Toby, Ashley Montanaro, and Stephen Piddock. "Universal Quantum Hamiltonians." arXiv preprint arXiv:1701.05182 (2017).

[90] Prokof'ev, N. V., B. V. Svistunov, and I. S. Tupitsyn. "“Worm” algorithm in

quantum Monte Carlo simulations." Physics Letters A 238.4 (1998): 253-257.
[91] Werner, Philipp, et al. "Continuous-time solver for quantum impurity models." Physical Review Letters 97.7 (2006): 076405.
[92] Troyer, Matthias, et al. "Non-local Updates for Quantum Monte Carlo Simulations." AIP Conference Proceedings. Vol. 690. No. 1. AIP, 2003.
[93] Sandvik, A. W., R. R. P. Singh, and D. K. Campbell. "Quantum Monte Carlo in the interaction representation: Application to a spin-Peierls model." Physical Review B 56.22 (1997): 14510.
[94] Swendsen, Robert H., and Jian-Sheng Wang. "Nonuniversal critical dynamics in Monte Carlo simulations." Physical review letters 58.2 (1987): 86.
[95] Sandvik, Anders W. "Stochastic series expansion method for quantum Ising models with arbitrary interactions." Physical Review E 68.5 (2003): 056701.
[96] Manousiouthakis, Vasilios I., and Michael W. Deem. "Strict detailed balance is unnecessary in Monte Carlo simulation." The Journal of chemical physics 110.6 (1999): 2753-2756.
[97] Liu, Cheng-Wei, Anatoli Polkovnikov, and Anders W. Sandvik. "Quantum versus classical annealing: insights from scaling theory and results for spin glasses on 3-regular graphs." Physical review letters 114.14 (2015): 147203.
[98] Morita, Satoshi, and Hidetoshi Nishimori. "Mathematical foundation of quantum annealing." Journal of Mathematical Physics 49.12 (2008): 125210.
[99] Hukushima, Koji, and Koji Nemoto. "Exchange Monte Carlo method and application to spin glass simulations." Journal of the Physical Society of Japan 65.6 (1996): 1604-1608.
[100] Machta, Jon. "Strengths and weaknesses of parallel tempering." Physical Review E 80.5 (2009): 056706.
[101] Farhi, Edward, et al. "Performance of the quantum adiabatic algorithm on random instances of two optimization problems on regular hypergraphs." Physical Review A 86.5 (2012): 052334.
[102] Hen, Itay, and A. P. Young. "Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems." Physical Review E 84.6 (2011): 061152.
[103] Jarret, Michael, Stephen P. Jordan, and Brad Lackey. "Adiabatic optimization versus diffusion Monte Carlo methods." Physical Review A 94.4 (2016): 042318.
[104] Bravyi, Sergey, et al. "The complexity of stoquastic local Hamiltonian problems." Quantum Inf. Comput., 8 (2008), pp. 0361.
[105] Bravyi, Sergey, and Barbara Terhal. "Complexity of stoquastic frustration-free Hamiltonians." Siam journal on computing 39.4 (2009): 1462-1485.
[106] Hastings, Matthew B., and M. H. Freedman. "Obstructions to classically simulating the quantum adiabatic algorithm." arXiv preprint arXiv:1302.5733 (2013).
[107] Hukushima, K., and Y. Iba. "Population annealing and its application to a spin glass." AIP Conference Proceedings. Vol. 690. No. 1. AIP, 2003.
[108] Katzgraber, Helmut G., et al. "Feedback-optimized parallel tempering Monte Carlo." Journal of Statistical Mechanics: Theory and Experiment 2006.03 (2006): P03018.
[109] Takahashi, Jun, Satoshi Takabe, and Koji Hukushima. "An Exact Algorithm Exhibiting RS-RSB/Easy-Hard Correspondence for the Maximum Independent Set Problem." Journal of the Physical Society of Japan 86.7 (2017): 073001.
[110] Karp, Richard M. "Reducibility among combinatorial problems." Complexity of computer computations. springer US, 1972. 85-103.
[111] M. Mézard, G. Parisi, and M. Á. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[112] M. Mézard, A. Montanari, *Information, Physics and Computation* (Oxford University Press, United States, 2009).

[113] Monasson, Rémi, et al. "Determining computational complexity from characteristic ' phase transitions ' ." Nature 400.6740 (1999): 133-137.

[114] Mulet, Roberto, et al. "Coloring random graphs." Physical review letters 89.26 (2002): 268701.

[115] Weigt, Martin, and Alexander K. Hartmann. "Minimal vertex covers on finite-connectivity random graphs: A hard-sphere lattice-gas picture." Physical Review E 63.5 (2001): 056127.

[116] Krzakala, Florent, et al. "Gibbs states and the set of solutions of random constraint satisfaction problems." Proceedings of the National Academy of Sciences 104.25 (2007): 10318-10323.

[117] Parisi, Giorgio. "Infinite number of order parameters for spin-glasses." Physical Review Letters 43.23 (1979): 1754.

[118] Edwards, Samuel Frederick, and Phil W. Anderson. "Theory of spin glasses." Journal of Physics F: Metal Physics 5.5 (1975): 965.

[119] Krzakala, F., and O. C. Martin. "Spin and link overlaps in three-dimensional spin glasses." Physical Review Letters 85.14 (2000): 3013.

[120] Barbara, B., A. P. Malozemoff, and Y. Imry. "Scaling of nonlinear susceptibility in MnCu and GdAl spin-glasses." Physical Review Letters 47.25 (1981): 1852.

[121] Kinzel, W., and K. Binder. "Static and dynamic magnetic response of spin-glass models with short-range interactions." Physical Review B 29.3 (1984): 1300.

[122] Morais, C. V., et al. "Spin-glass phase transition and behavior of nonlinear susceptibility in the Sherrington-Kirkpatrick model with random fields." Physical Review B 93.22 (2016): 224206.

[123] Marino, Raffaele, Giorgio Parisi, and Federico Ricci-Tersenghi. "The backtracking survey propagation algorithm for solving random K-SAT problems." Nature communications 7 (2016): 12996.

[124] Achlioptas, Dimitris, and Amin Coja-Oghlan. "Algorithmic barriers from phase transitions." Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on. IEEE, 2008.

[125] Cocco, Simona, and Rémi Monasson. "Statistical physics analysis of the computational complexity of solving random satisfiability problems using backtrack algorithms." The European Physical Journal B-Condensed Matter and Complex Systems 22.4 (2001): 505-531.

[126] Dubois, Olivier, and Jacques Mandler. "The 3-XORSAT threshold." Comptes Rendus Mathematique 335.11 (2002): 963-966.

[127] Cocco, Simona, et al. "Rigorous decimation-based construction of ground pure states for spin-glass models on random lattices." Physical review letters 90.4 (2003): 047205.

[128] Davis, Martin, and Hilary Putnam. "A computing procedure for quantification theory." Journal of the ACM (JACM) 7.3 (1960): 201-215. (2000).

[129] Weigt, Martin, and Alexander K. Hartmann. "Number of guards needed by a museum: A phase transition in vertex covering of random graphs." Physical review letters 84.26 (2000): 6118.

[130] Zhou, Haijun. "Vertex cover problem studied by cavity method: Analytics and population dynamics." The European Physical Journal B-Condensed Matter and Complex Systems 32.2 (2003): 265-270. Foundations of Computer Science, 364 (1981).

[131] Karp, Richard M., and Michael Sipser. "Maximum matching in sparse random graphs." Foundations of Computer Science, 1981. SFCS'81. 22nd Annual Symposium on. IEEE, 1981.

[132] Bauer, M., and O. Golinelli. "Core percolation in random graphs: a critical phenomena analysis." The European Physical Journal B-Condensed Matter and Complex Systems 24.3 (2001): 339-352.
[133] Takabe, Satoshi, and Koji Hukushima. "Typical behavior of the linear programming method for combinatorial optimization problems: a statistical-mechanical perspective." Journal of the Physical Society of Japan 83.4 (2014): 043801.
[134] Bender, Edward A., and E. Rodney Canfield. "The asymptotic number of labeled graphs with given degree sequences." Journal of Combinatorial Theory, Series A 24.3 (1978): 296-307.
[135] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." science 286.5439 (1999): 509-512.
[136] Takabe, Satoshi, and Koji Hukushima. "Typical performance of approximation algorithms for NP-hard problems." Journal of Statistical Mechanics: Theory and Experiment 2016.11 (2016): 113401.
[137] Bellman, Richard. The theory of dynamic programming. No. RAND-P-550. RAND CORP SANTA MONICA CA, 1954.
[138] Brélaz, Daniel. "New methods to color the vertices of a graph." Communications of the ACM 22.4 (1979): 251-256.
[139] Banderier, Cyril, et al. "Average case analysis of NP-complete problems: Maximum independent set and exhaustive search algorithms." Proceedings of AofA 9 (2009).
[140] Achlioptas, Dimitris, and Cristopher Moore. "Almost all graphs with average degree 4 are 3-colorable." Journal of Computer and System Sciences 67.2 (2003): 441-471.
[141] Raymond, Jack, Andrea Sportiello, and Lenka Zdeborová. "Phase diagram of the 1-in-3 satisfiability problem." Physical Review E 76.1 (2007): 011101.
[142] Parisi, Giorgio. "A sequence of approximated solutions to the SK model for spin glasses." Journal of Physics A: Mathematical and General 13.4 (1980): L115.
[143] Crosson, Elizabeth, and Aram W. Harrow. "Simulated quantum annealing can be exponentially faster than classical simulated annealing." Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on. IEEE, 2016.
[144] Kato, Tosio. "On the adiabatic theorem of quantum mechanics." Journal of the Physical Society of Japan 5.6 (1950): 435-439.
[145] Albash, Tameem, and Daniel A. Lidar. "Adiabatic quantum computing." arXiv preprint arXiv:1611.04471 (2016).
[146] Farhi, Edward, et al. "A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem." Science 292.5516 (2001): 472-475.
[147] Bapst, Victor, et al. "The quantum adiabatic algorithm applied to random optimization problems: The quantum spin glass perspective." Physics Reports 523.3 (2013): 127-205.
[148] Bapst, Victor, and Guilhem Semerjian. "On quantum mean-field models and their quantum annealing." Journal of Statistical Mechanics: Theory and Experiment 2012.06 (2012): P06007.
[149] Goldschmidt, Yadin Y. "Solvable model of the quantum spin glass in a transverse field." Physical Review B 41.7 (1990): 4858.
[150] Knysh, Sergey. "Zero-temperature quantum annealing bottlenecks in the spin-glass phase." Nature communications 7 (2016).
[151] Amin, M. H. S., and V. Choi. "First-order quantum phase transition in adiabatic quantum computation." Physical Review A 80.6 (2009): 062326.
[152] Altshuler, Boris, Hari Krovi, and Jérémie Roland. "Anderson localization makes adiabatic quantum optimization fail." Proceedings of the National Academy of Sciences 107.28 (2010): 12446-12450.
[153] Anderson, Philip W. "Absence of diffusion in certain random lattices." A Career In

Theoretical Physics. 2004. 79-93.

[154] Knysh, Sergey, 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).

[155] Ray, Pulak, Bikas K. Chakrabarti, and Arunava Chakrabarti. "Sherrington-Kirkpatrick model in a transverse field: Absence of replica symmetry breaking due to quantum fluctuations." Physical Review B 39.16 (1989): 11828.

[156] You, Wen-Long, Ying-Wai Li, and Shi-Jian Gu. "Fidelity, dynamic structure factor, and susceptibility in critical phenomena." Physical Review E 76.2 (2007): 022101.

[157] Gu, Shi-Jian. "Fidelity approach to quantum phase transitions." International Journal of Modern Physics B 24.23 (2010): 4371-4458.

[158] Garnerone, Silvano, et al. "Fidelity Approach to the Disordered Quantum X Y Model." Physical review letters 102.5 (2009): 057205.

[159] Fujiwara, Akio, and Hiroshi Nagaoka. "Quantum Fisher metric and estimation for pure state models." Physics letters A 201.2 (1995)119.

[160] Erdös, Paul, and Alfréd Rényi. "On the evolution of random graphs." Publ. Math. Inst. Hung. Acad. Sci 5.1 (1960): 17-60.

[161] Trotter, Hale F. "On the product of semi-groups of operators." Proceedings of the American Mathematical Society 10.4 (1959): 545-551.

[162] Sandvik, A. W. "A generalization of Handscomb's quantum Monte Carlo scheme-application to the 1D Hubbard model." Journal of Physics A: Mathematical and General 25.13 (1992): 3667.

[163] Wang, Lei, et al. "Fidelity susceptibility made simple: A unified quantum Monte Carlo approach." Physical Review X 5.3 (2015): 031007.

[164] Yasuda, Shinya, Hidemaro Suwa, and Synge Todo. "Stochastic approximation of dynamical exponent at quantum critical point." Physical Review B 92.10 (2015): 104411.

[165] Takahashi, Jun. "NP 完全問題の断熱量子計算における非一様な横磁場の効果". 修士論文. 東京大学大学院総合文化研究科広域科学専攻 (2015).

[166] Stein, Daniel L., and Charles Michael Newman. Spin glasses and complexity. Princeton University Press, 2013.

[167] Rényi, Alfréd. "On measures of entropy and information." Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California, 1961.

[168] Franchini, Fabio, A. R. Its, and V. E. Korepin. "Renyi entropy of the XY spin chain." Journal of Physics A: Mathematical and Theoretical 41.2 (2007): 025302.

[169] Žnidarič, Marko, Tomaž Prosen, and Peter Prelovšek. "Many-body localization in the Heisenberg X X Z magnet in a random field." Physical Review B 77.6 (2008): 064426.

[170] Takahashi, Kazutaka, and Yoshiki Matsuda. "Effect of random fluctuations on quantum spin-glass transitions at zero temperature." Journal of the Physical Society of Japan 79.4 (2010): 043712.

[171] Shitara, Tomohiro, and Masahito Ueda. "Determining the continuous family of quantum Fisher information from linear-response theory." Physical Review A 94.6 (2016): 062316.

[172] Mossi, Gianni, et al. "On the quantum spin glass transition on the Bethe lattice." Journal of Statistical Mechanics: Theory and Experiment 2017.1 (2017): 013102.

[173] Laumann, Christopher R., A. Pal, and A. Scardicchio. "Many-body mobility edge in a mean-field quantum spin glass." Physical review letters 113.20 (2014): 200405.

[174] Amin, Mohammad H. “Searching for quantum speedup in quasistatic quantum annealers.” Physical Review A 92.5 (2015): 052323.

[175] Goldschmidt, Yadin Y., and Pik-Yin Lai. “Ising spin glass in a transverse field: Replica-symmetry-breaking solution.” Physical review letters 64.21 (1990): 2467.

[176] Venuti, Lorenzo Campos, and Paolo Zanardi. “Quantum critical scaling of the geometric tensors.” Physical review letters 99.9 (2007): 095701.

[177] Gu, Shi-Jian. “Fidelity susceptibility and quantum adiabatic condition in thermodynamic limits.” Physical Review E 79.6 (2009): 061125.

[178] Selman, Bart, and Scott Kirkpatrick. “Critical behavior in the computational cost of satisfiability testing.” Artificial Intelligence 81.1 (1996): 273-295.

[179] Tao, Terence. “Finite time blowup for an averaged three-dimensional Navier-Stokes equation.” Journal of the American Mathematical Society 29.3 (2016): 601-674.